New Insights on Measuring Bank Branches Efficiency through DEA: Transactional, Operational, and Profit Assessments

Maria Conceição Andrade Silva Portela

Doctor of Philosophy

ASTON UNIVERSITY

October 2003

Summary

Aston University

New Insights on Measuring Bank Branches Efficiency through DEA: Transactional, Operational, and Profit Assessments

Maria Conceição Andrade Silva Portela

Doctor of Philosophy

2003

This thesis presents a number of methodological developments that were raised by a real life application to measuring the efficiency of bank branches.

The advent of internet banking and phone banking is changing the role of bank branches from a predominantly transaction-based one to a sales-oriented role. This fact requires the development of new forms of assessing and comparing branches of a bank. In addition, performance assessment models must also take into account the fact that bank branches are service and for-profit organisations to which providing adequate service quality as well as being profitable are crucial objectives.

This study analyses bank branches performance in their new roles in three different areas: their effectiveness in fostering the use of new transaction channels such as the internet and the telephone (transactional efficiency); their effectiveness in increasing sales and their customer base (operational efficiency); and their effectiveness in generating profits without compromising the quality of service (profit efficiency). The chosen methodology for the overall analysis is Data Envelopment Analysis (DEA).

The application attempted here required some adaptations to existing DEA models and indeed some new models so that some specialities of our data could be handled. These concern the development of models that can account for negative data, the development of models to measure profit efficiency, and the development of models that yield production units with targets that are nearer to their observed levels than targets yielded by traditional DEA models.

The application of the developed models to a sample of Portuguese bank branches allowed their classification according to the three performance dimensions (transactional, operational and profit efficiency). It also provided useful insights to bank managers regarding how bank branches compare between themselves in terms of their performance, and how, in general, the three performance dimensions are connected between themselves.

Keywords: Data Envelopment Analysis, Bank Branches, Negative Data, Profit Efficiency

Acknowledgments

Many were those that helped, supported, and contributed for the final result of this thesis. Without them this thesis and myself would not surely be the same!

My first thanks goes to my main supervisor Professor Emmanuel Thanassoulis, whose guidance was the driving force of all my work. Thanks for the questions you made that rose new topics to analyse and new routes to explore. Thanks for the time and patience you had with me when I was a bit stubborn, and thanks for all the encouragement that you gave me in all your feedbacks to my work.

Without any specific order after this, I would like to thank also to my other supervisors Professor Gary Simpson and Professor Rui Alves for all their support and useful comments on this work. I wish to thank especially to Professor Rui Alves for all his guidance and advice since I started teaching at the Catholic University, which lead me here today! Thanks also to Professor Elvira Silva for all her useful comments especially concerning the Chapter on profit efficiency where production economics (her area of expertise) played an important role.

A very special thanks to Pedro Borges for his friendship and for listening to some of the problems related with this work. We did some joint work on a paper that is also a Chapter of this thesis. His work on this paper is a valuable contribution to this thesis for which I cannot forget to heartily thank him. Thank you Pedro also for introducing me to the world of Latex without which the printed quality of this document would surely be worse. Thanks also for all the encouragement and good advice you gave me over the past few years!

To the Portuguese Catholic University my very special thanks for providing me with all the logistics and resources required to make a thesis as an external student in a foreign university. I wish to thank especially to Professor Alberto Castro (the director of the faculty of economics and management of Catholic University) for his support and encouragement during all the time I have spent teaching at this University and in particular during this

PhD work.

I would also like to acknowledge the financial support of the FCT (The Portuguese science and technology foundation), without which this thesis could not become a reality. I hope research funding in Portugal continues for many years to come.

To all the people with whom I contacted in the bank my heartily thanks! Unfortunately because the bank under analysis wishes to remain anonymous I cannot state any names here. Nevertheless, I would like to thank especially to the General Director (Director Geral Adjunto) for his help, kindness, time, and useful comments in all the meetings we had. Without his support I would have much more difficulties in gathering data and in getting other people from the bank to talk to me. Thanks so much for all that!

I also wish to thank to some colleagues both at the Aston University and at the Portuguese Catholic University for their precious help and support. To Camilo Valverde for his support in my first year assignments which he kindly read and gave me some valuable feedback. To Sandra Pereira for patiently listening to some of my problems and for her friendship and support since our first years of teaching at the Catholic University. To Sofia Salgado for her kind support and provision of some references relating to Service Quality, and to Ana Lourenço and Susana Lorga for their friendship during the last few years. To Leonardo Costa for helping me in my first steps in GAMS and to Pedro Duarte Silva for helping me with LIMDEP and some statistical issues. To Thoralf Dasslert for helping me sorting out some logistic problems with my installation in Aston and for his friendship for the last three years, and to Samma Athia for all her kindness and friendship during the time I spent at Aston. A very special thanks to Pam Lewis the secretary of the Aston Business school whose help was priceless, especially when I was not there to sort out various issues.

I also want to thank Ana Camanho for all the useful discussions we had around my work and hers (which was also related to assessing the efficiency of Portuguese bank branches), especially for some useful tips concerning the use of GAMS. To Susana Faria and Ana Viana my sincere thanks for their precious friendship during the past few years and for listening to my PhD problems during our regular lunches.

Finally I would like to thank to all my family to whom I heartily dedicate this thesis. Thanks Paulo for all your encouragement, love, support, and for keeping things going while I was in England. Thanks Patricia and Paulo for all your love and friendship, and Thanks Mother and Father for all your support and love during all my life. This thesis is for you!

I cannot finish my acknowledgments without dedicating this thesis also to a new person

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This work is the result of an interesting journey into two fields; one that relates to a recent methodology (Data Envelopment Analysis - DEA) that aims at analysing the efficiency of comparable production/decision units, and the other that relates to a deeper knowledge of these decision units, bank branches in our case.

The initial idea for this work was to compare two networks of bank branches in terms of their efficiency, and analyse the extent to which some factors could explain efficiency differences (if any) between the two networks of bank branches. It was, therefore, to be mainly an empirical work. As it is usual in PhD thesis the outcome is quite different from that initially planned and so it happened in this case. Indeed, the richness of the banking field was so big that soon it became clear that theoretical developments were also in need for the particular case of assessing the efficiency of bank branches. The first theoretical development that was in need was the creation of a model to assess the efficiency of bank branches that could account for their most recent challenges created by the increasing use of alternative distribution channels that threaten the survival of bank branches. Our research issue changed, therefore, to an analysis of the efficiency of a sample of bank branches that could account simultaneously for their changing role from transactional centers to retail centers, and for the objectives of bank branches as seen by the top management of the bank.

The literature on assessing bank branches efficiency is not wide (we found only 40 published papers on this matter and some working papers) and mostly the changing role of bank branches from transactional based to sales based has not been accounted for. At the same time no focus is given in any study to the development of alternative distribution channels and to the role that this development may have in improving the operational efficiency of bank branches. In fact, as more transactions are performed in alternative distribution channels, personnel at the bank branch is left with more time that can be used for performing other operational activities mostly sales related. In this sense, to foster the use of new distribution channels is an important objective of bank branches so that they can in fact focus on their new role as selling centers.

The model developed for analysing bank branches efficiency in their new context is presented in Chapter 5. Before that we review the literature on banking assessments in general (Chapter 3) and on bank branch assessments in particular (Chapter 4). In these literature reviews we focus particularly on studies that have applied the same methodology as we do, and therefore we introduce the main concepts about DEA in Chapter 2.

As a result of the aforementioned analysis three main dimensions of efficiency came out as the most important given the changing role of bank branches and their new objectives: *Transactional Efficiency*, *Operational Efficiency*, and *Profit Efficiency*. Transactional efficiency intends to capture the extent to which general transactions are being performed on other means than the bank branch, operational efficiency intends to measure the extent to which bank branches are increasing sales and the customer base of the bank branch, and profit efficiency intends to capture the extent to which bank branches are managing their product mix in a way that maximises profit.

Some problems were encountered and solved when each of these assessments was put into practice. In the operational efficiency measurement, for example, some of our outputs could be negative and original DEA formulations cannot handle negative data. For this reason, we developed a new model in Chapter 6 that could account satisfactorily for this type of data. The developed model has some advantages over existing models since it results in an efficiency score that is similar in meaning to radial efficiency scores usually obtained when data are non-negative. At the same time we also explored issues relating with target setting under negative data, where in particular we analysed the closeness of the obtained targets.

On the other hand, our intention of measuring profit efficiency encountered some obstacles since there are not many empirical applications applying the profit efficiency concept in a DEA context. In particular, to the authors knowledge, there is no study to date analysing the profit efficiency of bank branches either through parametric or non-parametric methodologies. At the same time the DEA literature on profit efficiency is still emergent and some developments were possible on this field. The developments on profit efficiency are put forward in Chapter 8, and consist of a new measure to compute profit efficiency and to decompose it into its technical and allocative components. In this Chapter we also use a framework developed in Chapter 7 for providing decision units with closest targets. This is a matter of interest in the DEA field since DEA models do not always result in target values that require the minimum effort of production units in moving towards the efficient frontier. In Chapter 9 the approach developed in Chapter 8 is further extended to the computation of total factor productivity change and its decomposition into a technical change, an efficiency change, and a residual component.

Using the foregoing theoretical developments we were able to measure the efficiency of a set of bank branches. This was done for the three dimensions of performance identified and for a time period that goes from March 2001 to September 2002 for the operational and profit assessments, and from January 2002 to September 2002 for the transactional assessment. The detailed results from these three assessments are presented in Chapter 10. Obviously the three aspects of performance are not necessarily independent as there may exist trade-offs between them. That is, it might be easily accepted that a bank branch is a good performer in operational terms, but not so good in profit terms. Our aim is to provide information on how well a bank branch is doing on each measure of performance, and cross-analyse results from all performance dimensions considered. This is done in Chapter 11, where we also analyse the relationship between efficiency and other variables such as location, age, competition, and service quality. Being banks and bank branches service organisations it results that the measurement of efficiency cannot be disentangled from that of service quality as detailed in Chapter 3. For this reason the analysis of the relationships between our performance dimensions and service quality assumed particular emphasis on Chapter 11. In Chapter 12 we conclude this thesis, pointing out its main methodological contributions, its main empirical results, and directions for further resaerch.

Chapter 2

Efficiency Measurement Through Data Envelopment Analysis

The main concept that will be applied throughout this study is the concept of efficiency. It is therefore important to explain what we mean by efficiency and how this concept is different and compares with others, such as effectiveness and productivity. In this Chapter we start by presenting these concepts and then we put forward usual procedures for efficiency measurement. Some concepts that will be referred throughout this thesis relating to efficiency measurement are also briefly described and explained in this Chapter.

2.1 Efficiency, Productivity, and Effectiveness

The concepts of efficiency and productivity are sometimes, but wrongly, taken as a single concept. We define **Efficiency** of a production unit as "the ratio of observed to maximum potential output obtainable from the given input, or the ratio of the minimum potential to observed input required to produce the given output, or some combination of the two" (Lovell, 1993, p. 4). The first ratio mentioned focus on outputs (it is output oriented) and the second focus on inputs (it is input oriented). The efficiency measure obtained from such ratios depends on the 'optimal' values taken as reference. If the optimal relates to the production function the resulting efficiency measure is technical, if it relates to any economic (cost, revenue or profit) function the resulting efficiency measure is economic.

Productivity, on the other hand, can be defined as a measure of the relationship between the outputs produced and the inputs used by a production unit.

To distinguish between the concepts of efficiency and productivity we use Figure 2.1

(this figure is identical to that shown in Coelli et al. (1998, p. 5)). In this figure $0F$

Figure 2.1: Productivity vs Efficiency

represents the production frontier defining the relationship between inputs (x) and outputs (y) for a particular industry. Any production unit operating on the production frontier (such as B or C) is technically efficient. If a production unit operates beneath the frontier (as it is the case of A) it is said to be inefficient. That is, such a unit could have produced more output given the input employed, or, it could have used less input given the output produced.

The ratio of the outputs produced to the inputs used (productivity ratio) is, for each unit, given by the slope of the lines departing from the origin and intercepting each point (A, B, and C). At point C the slope of the ray from the origin is the highest, meaning that this point defines the maximum possible productivity for this industry. Moving from point B to point C represents an exploitation of scale economies, as point C is the point of optimal scale, but implies no gains in efficiency. We can thus conclude that an efficient production unit may still be able to improve its productivity by exploiting scale economies (Coelli et al., 1998).

Concerning the productivity concept it is still important to distinguish between total factor productivity and partial measures of productivity (Coelli et al., 1998). The former involves all factors of production, and the latter focus only on some production factors (a commonly used example is labour productivity). Note that calculating aggregate measures of productivity (a ratio of aggregate output per aggregate input) implies the definition of an aggregation formula like a simple or weighted average. Traditional index number approaches (e.g. Py, 1990) provide such a way, where factor's prices are the aggregating factor. The best well known examples of such indexes are the Laspeyres, Paasche and Fisher Indexes. Without price information the aggregation of amounts expressed in different units of measurement, is not free of controversy as will become clear in subsequent chapters.

Effectiveness can be defined as the extent to which an organisation meets its objectives or goals (Klassen et al., 1998; Agrell and Bogetoft, 2001). In this sense effectiveness can be measured as the ratio between produced outputs and desired outputs. The most common distinction between efficiency and effectiveness is that the former is 'to do things right' while the latter is 'to do the right thing'. Or, in the words of Golany et al. (1993), efficiency is related to *performing current activities as well as possible* and effectiveness is related to *choosing the proper activities.* Effectiveness is therefore linked with strategic choices and strategic goals whose attainability is desired.

The above definitions intend to establish at the beginning of this study what is understood by the terms efficiency, productivity and effectiveness. Note that the above definitions are consistent with what is currently accepted amongst economists and Management Science/Operational Research scientists. This does not exclude the existence of other understandings of the above concepts in different scientific areas, as the works of Achabal et al. (1984) and Klassen et al. (1998) testify.

2.1.1 Some Thoughts around and about Efficiency

From the three concepts defined in the previous section, efficiency is the core concept of this study. In the context of measuring efficiency one can accept as given the fact that firms may be inefficient, or one can raise some doubts about the meaning of measuring efficiency. In the latter case questions like *what does efficiency really mean?*; or *what makes a firm to produce less than the maximum possible quantity of outputs?* are inevitable. "While there has been a proliferation of sophisticated techniques for estimating the frontier production function, relatively little attention has been devoted to interpretation of measured inefficiency" (Ray, 1988, p. 167). This last issue is, however, as important as the first. Classic production economics assume that in the long run there are no inefficiencies. This statement is based on a number of assumptions, among which the assumption that the market is perfectly competititive[1]. Under this assumption, a comparison between two firms using the same input combinations but producing differing amounts of outputs leads to the conclusion that necessarily some relevant inputs that are not equal for both firms have been neglected in this comparison (Ray, 1988). Indeed, if we think of a situation where all the

[1]Perfect competition assumes a large number of production units, perfect mobility of factors, and perfect information.

factors by which firms can differ are accounted for in a comparison of these, then very few inefficiencies could remain. Under this perspective it is meaningless to measure efficiency (especially as far as technical efficiency is concerned), because all inefficiency is associated with *model mis-specification* reflecting variables that were not included in the efficiency assessment, or wrong assumptions concerning functional relationships between inputs and outputs (Bogetoft and Hougaard, 2001).

In our opinion the traditional definition of efficiency as 'doing things right' is the one that justifies the need to measure efficiency. Indeed, production units do not always do things right and more often than desired they do things wrongly. This is just a consequence of organisations being constituted by human beings and not by machines. The economic theory accords in a certain sense to a positivist approach that regards organisations as machines. In such a setting inefficiency would not take place. Going a little beyond that, albeit not necessarily abandoning a positivist approach, inefficiency is not only possible but also a stark reality. If one thinks of two bank branches operating exactly under the same environmental conditions and having access exactly to the same resources, would these necessarily operate in the same manner? would both be equally efficient? We believe the answer is no. For example one of the resources that the branch might have at its disposal is a computer program to manage its clients base. Depending on the motivation and dynamics of the people in the branch this program can be exploited in its full potential or not. Obviously, one can argue that in that sense the human resources of the two branches are not equal because one has more motivated and dynamic people than the other. While this is true it is also true that no two people are equal and the assumption of equal resources as far as humans are concerned is basically impossible to achieve. So in the limit efficiency is always related with people. It is people that are efficient or inefficient and not machines. Two equal machines are necessarily equally efficient. If machines do things in a wrong way, the inefficiency here is not attributed to the machine but to its programmer.

Contributing to the discussion around the concept and measurement of efficiency Bogetoft and Hougaard (2001) put forward a novel interpretation of efficiency according to which inefficiency is a *rational choice*. That is, some decision units choose to be inefficient because the costs of increasing efficiency would be higher than the costs of remaining inefficient. This is undoubtedly an interesting interpretation, as most of the times the costs of adjusting efficiency are not accounted for in efficiency assessments. If these were accounted for, then in some situations one could indeed conclude that it is cheaper to stay inefficient rather than to try to improve efficiency.

We will take the perspective in this study that it is possible to measure inefficiency because inefficiency is a *waste* (Bogetoft and Hougaard, 2001). That is, inefficient firms employ too many inputs for producing a certain level of outputs, or produce too few outputs from a given set of inputs. The 'optimal' level of inputs or outputs to which every production unit compares its own input/output levels to assess its efficiency lie on a production frontier (for technical efficiency measurement). This production frontier can be broadly of two types: deterministic or stochastic (see Førsund et al., 1980, for a complete characterisation of various approaches). Deterministic frontiers assume that, for a given level of input, there is an exact value of maximum output that is possible, while stochastic frontiers assume that the maximum output for a given level of input is random rather than exact. In both cases the interest is to measure the distance of each observation to the frontier, but in stochastic frontiers the form of that frontier is an *assumption* of all the analysis while in deterministic frontiers the form of the frontier is a *result* of the analysis. Therefore these two broad classes of methods begin the analysis in different and opposite directions, albeit being interested in obtaining the same type of results.

This study uses a particular class of deterministic frontiers best known as data envelopment analysis (DEA) after the developments of Charnes et al. (1978). Under this class of methods there are two decisions to be made before proceeding with the efficiency measurement exercise. The first concerns the choice of the technological set (bounded by a production frontier) against which efficiency will be measured, and the second concerns the choice of the efficiency measure to use. These issues will be addressed in the next sections.

2.2 Production Technology

Consider an input vector $\mathbf{x} = (x_1, \ldots, x_m) \in R_+^m$ used to produce an output vector $\mathbf{y} = (y_1, \ldots, y_s) \in R_+^s$ in a technology involving n production units. The production possibilities set, T, describes all patterns of inputs and outputs that are technologically feasible (Varian, 1992). T is thus the production technology that transforms $\mathbf{x}$ in $\mathbf{y}$, and is defined as:

$$T = \{(\mathbf{x}, \mathbf{y}) \in R_+^{m+s} \mid \mathbf{x} \text{ can produce } \mathbf{y}\} \tag{2.1}$$

The production technology T, or graph technology, may also be represented from two other perspectives: input or consumption set and production or output set. An input set

$L(\mathbf{y})$ is the subset of all input vectors $\mathbf{x} \in R_+^m$ yielding at least $\mathbf{y}$, and a production set $P(\mathbf{x})$ is the subset of all output vectors $\mathbf{y} \in R_+^s$ which are obtained from $\mathbf{x}$.

$$\begin{aligned} &L(\mathbf{y}) = \{\mathbf{x} \mid (\mathbf{x},\mathbf{y}) \in T\} \text{ or } L(\mathbf{y}) = \{\mathbf{x} \mid \mathbf{y} \in P(\mathbf{x})\} \\ &\text{and} \\ &P(\mathbf{x}) = \{\mathbf{y} \mid (\mathbf{x},\mathbf{y}) \in T\} \text{ or } P(\mathbf{x}) = \{\mathbf{y} \mid \mathbf{x} \in L(\mathbf{y})\} \end{aligned} \tag{2.2}$$

A production technology can have, therefore, alternative and equivalent representations highlighting different aspects of that technology. "The input set models input substitution, and the output set models output substitution. The graph models both input substitution and output substitution, in addition to modelling input-output transformation" (Färe et al., 1994a, p. 27).

A production technology defined by $L(\mathbf{y})$, $P(\mathbf{x})$ or $T(\mathbf{x},\mathbf{y})$ has some relevant subsets that are useful for efficiency measurement. The characteristics of these correspondences as well as the analysis of these subsets is detailed in Färe et al. (1985). For our purposes we are mainly interested in two subsets: the isoquant and the efficient subset.

- The input isoquant of $L(\mathbf{y})$, the output isoquant of $P(\mathbf{x})$, and the graph isoquant of $T(\mathbf{x},\mathbf{y})$ are defined as:
  $Isoq\ L(\mathbf{y}) = \{\mathbf{x} \mid \mathbf{x} \in L(\mathbf{y}), \lambda\, \mathbf{x} \notin L(\mathbf{y}), \lambda \in [0,1[\}$,
  $Isoq\ P(\mathbf{x}) = \{\mathbf{y} \mid \mathbf{y} \in P(\mathbf{x}), \theta\, \mathbf{y} \notin P(\mathbf{x}), \theta > 1\}$,
  $Isoq\ T(\mathbf{x},\mathbf{y}) = \{(\mathbf{x},\mathbf{y}) \mid (\mathbf{x},\mathbf{y}) \in T(\mathbf{x},\mathbf{y}), (\lambda\, \mathbf{x}, \lambda^{-1}\mathbf{y}) \notin T(\mathbf{x},\mathbf{y}), \lambda \in]0,1[\}$;

- The efficient subsets of $L(\mathbf{y})$, $P(\mathbf{x})$, and $T(\mathbf{x},\mathbf{y})$ are defined as:
  $Eff\ L(\mathbf{y}) = \{\mathbf{x} \mid \mathbf{x} \in L(\mathbf{y}), \mathbf{x}' \leq \mathbf{x} \text{ and } \mathbf{x}' \neq \mathbf{x} \ \Rightarrow \mathbf{x}' \notin L(\mathbf{y})\}$,
  $Eff\ P(\mathbf{x}) = \{\mathbf{y} \mid \mathbf{y} \in P(\mathbf{x}), \mathbf{y}' \geq \mathbf{y} \text{ and } \mathbf{y}' \neq \mathbf{y} \ \Rightarrow \mathbf{y}' \notin P(\mathbf{x})\}$,
  $Eff\, T(\mathbf{x},\mathbf{y}) = \{(\mathbf{x},\mathbf{y}) \mid (\mathbf{x},\mathbf{y}) \in T(\mathbf{x},\mathbf{y}), (-\mathbf{x}',\mathbf{y}') \geq (-\mathbf{x},\mathbf{y}) \text{ and } (-\mathbf{x}',\mathbf{y}') \neq (-\mathbf{x},\mathbf{y}) \Rightarrow (\mathbf{x}',\mathbf{y}') \notin T(\mathbf{x},\mathbf{y})\}$.

These definitions imply that $Isoq\ L(\mathbf{y}) \supseteq Eff\ L(\mathbf{y})$ (see Färe et al. (1985) for general conditions under which the two subsets coincide). The same relationship is valid for the output and graph cases.

We illustrate in Figures 2.2 and 2.3 the input and output correspondences, respectively. The efficient subset in both cases is the set of points on $Isoq\ L(\mathbf{y})$ or $Isoq\ P(\mathbf{x})$ that lie between the identified rays.

Figure 2.2: Input Space Representation

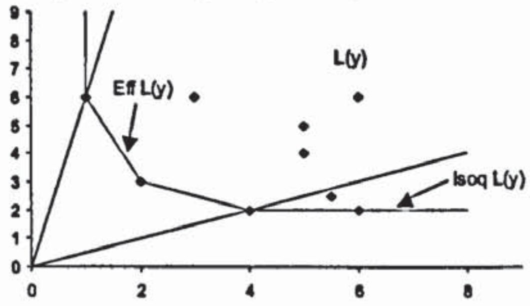
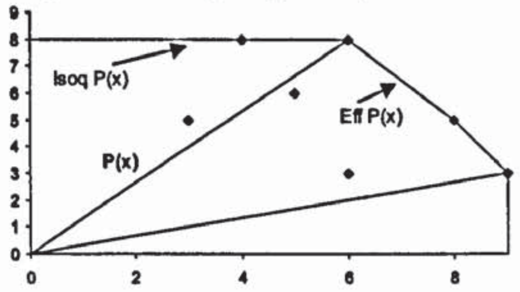


Figure 2.3: Output Space Representation



We will characterise technologies through piecewise linear mathematical formulae (e.g. Charnes et al., 1978; Färe et al., 1985) constructed from observed data. Consider the variables $(\lambda_1, \ldots, \lambda_j)$ denoting the intensity level of each of the j activities. The input reference set $L(\mathbf{y})$, satisfying strong disposability of inputs and outputs[2], is represented by (2.3)

$$L(\mathbf{y}) \quad = \quad \left\{\mathbf{x} \mid \sum_{j=1}^{n} \lambda_j \, \mathbf{y}_j \geq \mathbf{y}, \ \sum_{j=1}^{n} \lambda_j \, \mathbf{x}_j \leq \mathbf{x}, \ \lambda_j \ \geq \ 0, \ j \ = \ 1, \ldots, n\right\} \quad (2.3)$$

Equivalently the output reference set, satisfying strong disposability of inputs and outputs, is represented by (2.4). Note that strong (or free) disposability of all inputs and outputs is equivalent to monotonicity of the corresponding production set (e.g. Kuosmanen, 2001)).

$$P(\mathbf{x}) \quad = \quad \left\{\mathbf{y} \mid \sum_{j=1}^{n} \lambda_j \, \mathbf{y}_j \geq \mathbf{y}, \ \sum_{j=1}^{n} \lambda_j \, \mathbf{x}_j \leq \mathbf{x}, \ \lambda_j \ \geq \ 0, \ j \ = \ 1, \ldots, n\right\} \quad (2.4)$$

Depending on the properties satisfied by $L(\mathbf{y})$ and $P(\mathbf{x})$ its geometric form in Figures 2.2 and 2.3 could be different[3].

A production function describes quantitatively the technological relationship between outputs and inputs of a production process. It expresses the maximum output realisable

[2]Strong disposability of inputs implies that if a given quantity of inputs can produce a given amount of outputs, then any higher amount of inputs is also capable of producing the same amount of output. Strong disposability of outputs implies that if a given quantity of outputs is produced from a given level of inputs, then less outputs can always be produced from the same amount of inputs. Mathematically both these properties imply that for $\mathbf{z} = (-\mathbf{x}, \mathbf{y}) \in T$ if $\mathbf{z}' \leq \mathbf{z}$, then $\mathbf{z}' \in T$.

[3]For example if we considered weak disposability of inputs instead of strong disposability, the form of $L(\mathbf{y})$ would change. Weak disposability of inputs states that if $\mathbf{x} \in L(\mathbf{y})$, then $\lambda \, \mathbf{x} \in L(\mathbf{y})$ for $\lambda \geq 1$. This assumption considers that proportional increases in all inputs can yield the production of the same amount of output [see Färe et al. (1985) for details]. Weak disposability of outputs is defined in an analogous way.

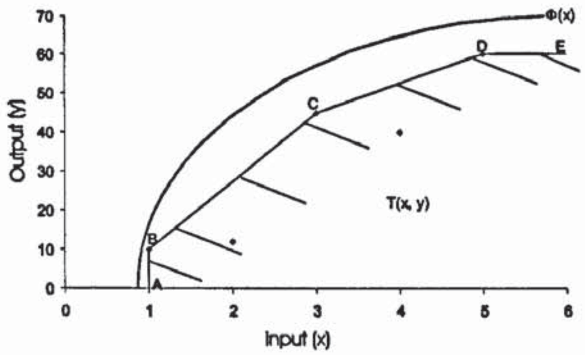
from a set of inputs, being thus the frontier of the set T.

The production function, or production frontier can be defined as:

$$\Phi(\mathbf{x}) = Max\{\mathbf{y} \mid (\mathbf{x}, \mathbf{y}) \in T\} \tag{2.5}$$

The theoretical production function (2.5) is by definition unknown. It can, however, be approximated by a piecewise linear technology as shown in Figure 2.4.

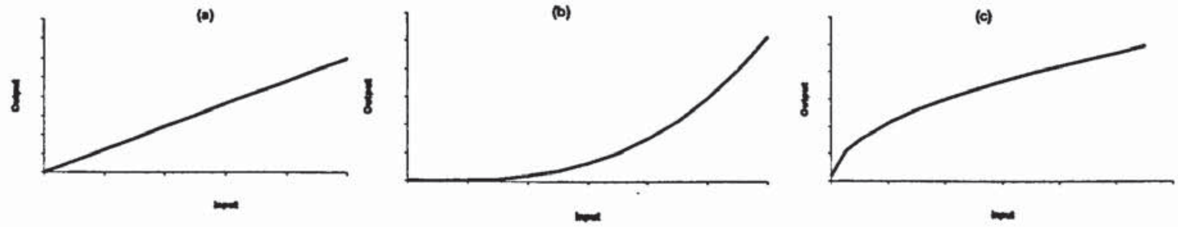
Figure 2.4: Production Function



The production technology $T(\mathbf{x}, \mathbf{y})$ represented in Figure 2.4 is bounded by ABCDE, which is also the isoquant of the production technology. Its efficient subset is represented by BCD, since segments BA and DE are not efficient in a Pareto-Koopmans sense. Pareto-Koopmans efficiency, or simply Pareto efficiency, is attained when an increase in any output (or a decrease in any input) requires a decrease in at least another output (or an increase in at least another input) (e.g. Lovell, 1993).

One of the characteristics of any production technology is returns to scale (RTS). One can define a technology exhibiting globally constant returns to scale (CRS) (Figure 2.5(a)), globally increasing returns to scale (IRS) (Figure 2.5(b)), or globally decreasing returns to scale (DRS) (Figure 2.5(c)). The piecewise linear representation of $T(\mathbf{x}, \mathbf{y})$ on Figure 2.4

Figure 2.5: Returns to Scale



does not satisfy globally any of the above RTS, meaning that it satisfies variable returns to scale (VRS). That is, returns to scale are increasing until point C and decreasing after point C. Point C has maximum productivity and satisfies locally constant returns to scale. The piecewise mathematical representation of VRS and CRS technologies based on observed values is shown in (2.6) and (2.7), respectively.

$$T^{VRS} = \left\{(\mathbf{x},\mathbf{y}) \mid \sum_{j=1}^{n}\lambda_j\, \mathbf{y}_j \geq \mathbf{y},\ \sum_{j=1}^{n}\lambda_j\, \mathbf{x}_j \leq \mathbf{x},\ \sum_{j=1}^{n}\lambda_j = 1,\ \lambda_j \geq 0,\ j = 1,\ldots,n\right\} \quad (2.6)$$

$$T^{CRS} = \left\{(\mathbf{x},\mathbf{y}) \mid \sum_{j=1}^{n}\lambda_j\, \mathbf{y}_j \geq \mathbf{y},\ \sum_{j=1}^{n}\lambda_j\, \mathbf{x}_j \leq \mathbf{x},\ \lambda_j \geq 0,\ j = 1,\ldots,n\right\} \quad (2.7)$$

A technological set exhibiting CRS assumes that replication is possible. That is, if inputs are scaled up or down by some amount γ it is assumed that the amount γ times the output produced before can also be produced (see Varian, 1992, p. 15).

The VRS and CRS technological sets were the first sets against which efficiency was measured through DEA. Charnes et al. (1978) measured efficiency in relation to CRS technologies only, and Banker et al. (1984) first measured efficiency in relation to VRS technologies. Both the CRS and VRS technological sets are convex[4]. The convexity of T implies that both the input and output sets [$L(\mathbf{y})$ and $P(\mathbf{x})$] are convex, although the inverse is not true (see Petersen, 1990). Note that convexity was assumed in our previous graphical representations of $L(\mathbf{y})$, $P(\mathbf{x})$, and $T(\mathbf{x},\mathbf{y})$.

Convexity is assumed in most economic models of production, but there is some debate in the literature concerning the need for this assumption. In fact assuming convexity implies that some returns to scale characteristics of a production set cannot be modelled. For example the technological set (b) depicted in Figure 2.5 representing global increasing returns to scale is not convex and therefore cannot be represented by any of the above piecewise linear representations of the technology. In fact the convexity of the production set excludes the possibility of globally increasing returns to scale and other possibilities where alternate behaviours of increasing and decreasing returns at different volumes take place (Bogetoft et al., 2000). At the same time convexity is also inconsistent with decreasing marginal rates of transformation for the outputs, and decreasing marginal rates of substitution for the inputs[5] (Post, 1999; Banker and Maindiratta, 1986). In situations where commodities

[4] T is a convex set if $\forall\, \mathbf{z} = (-\mathbf{x},\mathbf{y})$ and $\mathbf{z}' = (-\mathbf{x}',\mathbf{y}') \in T$, $\forall\, \alpha \in [0,1], \alpha\, \mathbf{z} + (1-\alpha)\, \mathbf{z}' \in T$.

[5] Marginal rate of substitution is the slope of the isoquant of $L(y)$. It reflects the rate at which one input

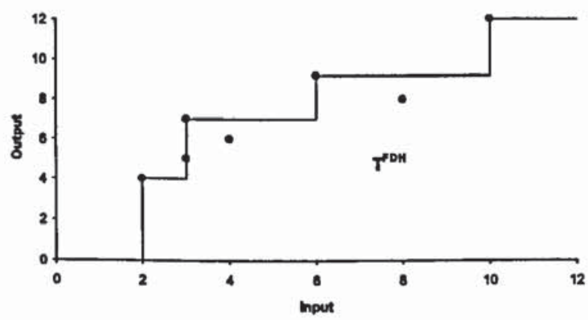
are not continuously divisible the assumption of convexity does not apply as well (Coelli et al., 1998). The main reasons for assuming convexity of T are, therefore, on the one hand the neoclassical assumption of diminishing marginal rates of substitution and, on the other hand, the fact that convexity is a necessary assumption for establishing the duality between input and output sets and cost and revenue functions (Petersen, 1990).

In order to model situations where the convexity of the technological set is not an adequate assumption, some non-convex production possibilities sets have been developed. The best known non-convex technological set that only satisfies free disposability of inputs and outputs is the Free Disposal Hull (FDH), which was first introduced by (Deprins et al., 1984). The production possibility set of this technology is defined in (2.8).

$$T^{FDH} = \left\{ (\mathrm{x},\mathrm{y}) \mid \sum_{j=1}^{n} \lambda_j \, \mathrm{y_j} \geq \mathrm{y},\ \sum_{j=1}^{n} \lambda_j \, \mathrm{x_j} \leq \mathrm{x},\ \sum_{j=1}^{n} \lambda_j = 1,\ \lambda_j \in \{0,1\},\ j = 1,\ldots,n \right\} \tag{2.8}$$

The particularity of a technology defined by T^{FDH} is that it rests only on the assumption of free disposability of inputs and outputs. The non convex nature of T^{FDH} is expressed on the binary constraints associated to the λ_j values. Graphically a FDH technology looks like that shown in Figure 2.6. An interesting characteristic of T^{FDH} is the fact that the efficient

Figure 2.6: FDH Production Set



subset of the production frontier is constituted by observable units only. This makes FDH a useful method to be applied for benchmarking purposes. As pointed out by Bogetoft et al. (2000, p. 2) "fictitious production possibilities, generated as convex combinations of those actually observed, are usually less convincing as benchmarks, or reference units, than actually observed production possibilities".

can be substituted by another while holding everything else constant. The marginal rate of transformation reflects the same relationship between outputs.

The three technological sets mentioned hitherto are the ones that will be most referred throughout this study. Other non-convex specifications of the technology can be found in Banker and Maindiratta (1986), Petersen (1990), Tulkens (1993), Bogetoft (1996), Kuosmanen (1999a), Bogetoft et al. (2000), Post (2001a), Post (2001b), Kuosmanen (2001), Agrell et al. (2001), and Kuosmanen and Post (2002).

2.3 Efficiency Measurement

After specifying the technology against which efficiency is to be measured one needs to decide on the type of efficiency measure to use. In this section we will detail on some known models to measure efficiency. In later chapters additional references to the issue of measuring efficiency will be made.

A measure of efficiency is intended to capture the distance from each observation to the technological frontier. Farrell (1957) defined a distance function (F_j) that is still the most used measure of efficiency in the literature. Such a measure is usually defined either on the input space or the output space. The Farrell input technical efficiency measure of observation $o \in L(\mathbf{y})$ can be defined as [see Färe and Lovell (1978) or Färe et al. (1985)]:

$$FI_o = \min\{\theta \mid \theta\mathbf{x} \in L(\mathbf{y})\}. \tag{2.9}$$

The Farrell output technical efficiency measure of observation $o \in P(\mathbf{x})$ can be defined as:

$$FO_o = \max\{\beta \mid \beta\mathbf{y} \in P(\mathbf{x})\} \tag{2.10}$$

The linear programming formulations that allow the computation of the efficiency measures θ and β are shown in models (2.11) and (2.12), respectively.

$$\begin{array}{llll}
\min\Big\{\theta_o \mid & & \max\Big\{\beta_o \mid & \\
\sum_{j=1}^{n}\lambda_j\, y_{rj} - s_r = y_{ro} & r = 1,\ldots,s & \sum_{j=1}^{n}\lambda_j\, y_{rj} - s_r = \beta_o y_{ro} & r = 1,\ldots,s \\
\sum_{j=1}^{n}\lambda_j\, x_{ij} + e_i = \theta_o\, x_{io} & i = 1,\ldots,m & \sum_{j=1}^{n}\lambda_j\, x_{ij} + e_i = x_{io} & i = 1,\ldots,m \\
\lambda_j \in S, \quad s_r,\ and\ e_i \geq 0\Big\} & (2.11) & \lambda_j \in S, \quad s_r,\ and\ e_i \geq 0\Big\} & (2.12)
\end{array}$$

Depending on S various technological sets can be defined especially as far as returns to scale characteristics are concerned. Originally Farrell considered only CRS technologies where $S = \{\lambda_j \geq 0\}$. However, other possibilities are:

- $S = \{\sum_{j=1}^{n} \lambda_j = 1\} \ \Rightarrow$ VRS;
- $S = \{\sum_{j=1}^{n} \lambda_j \leq 1\} \ \Rightarrow$ NIRS;
- $S = \{\sum_{j=1}^{n} \lambda_j \geq 1\} \ \Rightarrow$ NDRS;
- $S = \{\lambda_j \in \{0,1\}\} \ \Rightarrow$ FDH.

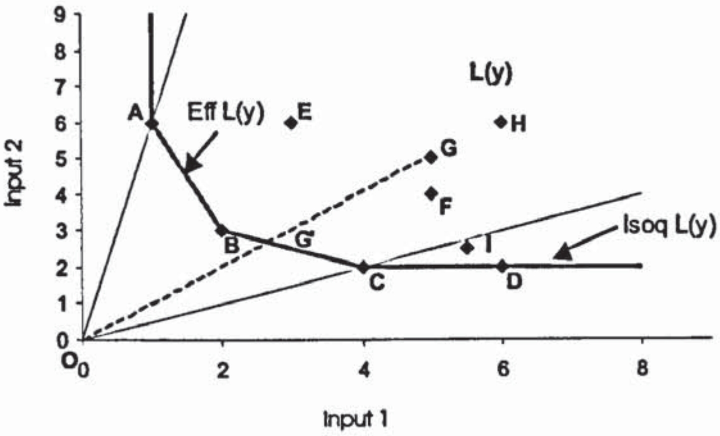
The CRS, VRS, and FDH technological sets were already referred to previously. The other two sets impose non increasing returns to scale (NIRS) and non decreasing returns to scale (NDRS).

The θ and β efficiency measures of a given unit o correspond to the ratio of an input or output vector located on the production frontier to the observed input or output vector. If the optimum value of theta is $\theta_o^* < 1$, then the unit is not technically efficient as it is possible to produce the same output vector with the reduced input vector $\theta_o^* \mathbf{x}$. If the optimum value of beta is $\beta_o^* > 1$, then the unit is not technically efficient as it is possible to use the same input vector to produce the augmented output vector $\beta_o^* \mathbf{y}$. When $\theta^* = 1$ or $\beta^* = 1$ the observed production unit lies on the frontier of the technological set and is deemed 100% efficient.

The reference technological set used in Farrell measures of efficiency is the isoquant of $L(\mathbf{y})$ or the isoquant of $P(\mathbf{x})$ and not its efficient subsets. This means that an optimal value of 1 for θ^* or β^* in models (2.11) and (2.12), respectively, does not necessarily accord with the Pareto-Koopmans notion of efficiency. Pareto-efficiency is assured only when $\theta^* = 1$ or $\beta^* = 1$, and all optimal slack values (s_r^* and e_i^*) are zero (Charnes et al., 1978). In an optimal solution to model (2.11) (or (2.12)), the value of θ^* (or β^*) represents equiproportional or radial changes in all inputs (or outputs) that project any observation on the technological frontier. The optimal value of slacks represents additional sources of inefficiency that are not accounted for by the radial θ^* (or β^*) factor.

Figure 2.7 illustrates input efficiency measurement for the case of inefficient unit G. Units A, B, C, and D, in Figure 2.7, are Farrell input technical efficient. Unit G is inefficient as it could produce the same amount of output using fewer of both its inputs. The radial distance of this unit to the input isoquant is given by $\theta_G = \frac{OG'}{OG}$. Note that unit D in Figure

Figure 2.7: Input Efficiency Measurement



2.7 lies on the isoquant but is not Pareto-efficient. In fact, unit D is able to reduce the usage of one input without changing the usage of the other.

The dual of models (2.11) and (2.12) are shown in (2.13) and (2.14), respectively, for the most general case where $S = \{\sum_{j=1}^{n} \lambda_j = 1\}$.

$$\max\Big\{h_o = \sum_{r=1}^{s} u_r\, y_{ro} + u_o \mid \sum_{i=1}^{m} v_i\, x_{io} = 1,\ \sum_{r=1}^{s} u_r\, y_{rj} - \sum_{i=1}^{m} v_i\, x_{ij} + u_o \leq 0,\ j = 1,\ldots,n,\ \ u_r, v_i \geq 0,\ u_o \text{ is free}\Big\} \quad (2.13)$$

$$\min\Big\{g_o = \sum_{i=1}^{m} v_i\, x_{io} + u_o \mid \sum_{r=1}^{s} u_r\, y_{ro} = 1,\ \sum_{i=1}^{m} v_i\, x_{ij} - \sum_{r=1}^{s} u_r\, y_{rj} + u_o \geq 0,\ j = 1,\ldots,n,\ \ u_r, v_i \geq 0,\ u_o \text{ is free}\Big\} \quad (2.14)$$

The dual of the Farrell efficiency measures can accommodate both VRS and CRS cases. To specify a VRS technology let u_o assume any value (positive, negative or zero), otherwise to specify a CRS technology set u_o to zero.

In models (2.13) and (2.14) the variables v_i and u_r represent the weights that the assessed unit allocates to each of its inputs and outputs so that its efficiency is maximised. The unit is free to choose these weights and as such the efficiency measure will show it in the best possible light.

Models of the type in (2.11) and (2.12) will be called envelopment models, while their duals in (2.13) and (2.14) will be called multiplier DEA models. The solution of an envelopment model results in the values of inputs and/or outputs that a unit should attain if efficient. Such values are usually called targets and are generally defined in (2.15).

$$
\begin{aligned}
y^*_{ro} &= \sum_{j=1}^{n} \lambda^*_j \, y_{rj} = \beta^*_o \, y_{ro} + s^*_r \qquad r = 1, \ldots, s \\
x^*_{io} &= \sum_{j=1}^{n} \lambda^*_j \, x_{ij} = \theta^*_o \, x_{io} - e^*_i \qquad i = 1, \ldots, m
\end{aligned} \tag{2.15}
$$

In input oriented models such as (2.11) β^*_o is assumed to be 1 in (2.15), and in output oriented models such as (2.12) θ^*_o is assumed to be 1. Target inputs and outputs are therefore the result of a linear combination of a set of units located on the efficient frontier for which λ is non-zero in the solution of model (2.11) or (2.12). Such units, used in the construction of target levels, are called the peer units of the unit o being assessed.

Note that the target levels in (2.15) are not radial targets (resulting from equiproportional contraction of inputs or outputs) because they include optimal slacks. Nevertheless, targets in (2.15) are not necessarily Pareto-efficient. Note for example that assessing unit I in Figure 2.7 renders target input levels equal to $(x^*_1, x^*_2) = (4.4, 2)$ located between units C and D, the peers of unit I. These target inputs do not lie on the efficient subset of $L(\mathbf{y})$. In order to guarantee that targets lie on the efficient subset of $L(\mathbf{y})$ one needs to solve a second stage model where the slacks are maximised and target levels in (2.15) are used on the right hand side of (2.16) (e.g. Ali and Seiford, 1993b).

$$
\begin{aligned}
&\max \Big\{ \sum_{r=1}^{s} s_r + \sum_{i=1}^{m} e_i \mid \\
&\sum_{j=1}^{n} \lambda_j \, y_{rj} - s_r = y^*_{ro} \qquad r = 1, \ldots, s \\
&\sum_{j=1}^{n} \lambda_j \, x_{ij} + e_i = \ x^*_{io} \qquad i = 1, \ldots, m \\
&\lambda_j \in S, \quad s_r, \ and \ e_i \geq 0 \Big\}
\end{aligned} \tag{2.16}
$$

If the optimal solution of this second stage model is zero (that is all slacks are zero) then the optimal target levels resulting from the first stage are Pareto-efficient. If the optimal solution of this model is different from zero then the second stage model solution provides Pareto-efficient targets and peers. For unit I in Figure 2.7, the second stage model identifies a value of $e^*_1 = 0.4$ and $\lambda^*_C = 1$. This means that Pareto-efficient targets for unit I are $(x^*_1, x^*_2) = (4, 2)$, being its only peer unit C.

The procedure for finding Pareto-efficient targets can be done through the second phase procedure outlined above, or it can be done in a single phase if the objective function of (2.11) or (2.12) is replaced by (2.17).

$$\theta_o - \epsilon(\sum_{r=1}^{s} s_r + \sum_{i=1}^{m} e_i) \qquad \text{or} \qquad \beta_o + \epsilon(\sum_{r=1}^{s} s_r + \sum_{i=1}^{m} e_i) \tag{2.17}$$

This specification was first proposed by Charnes et al. (1978), and it basically means that preemptive priority is given to the minimisation of θ, or to the maximisition of β, and secondly the maximisation of slacks is sought. Because slacks are multiplied by a very small value (identified by ϵ), the resulting objective function from (2.17) is in fact equal to that resulting from the radial model presented before. Although the single stage approach is theoretically correct, it may result in computational inaccuracies and erroneous results when the value of ϵ is specified (for details see Ali and Seiford, 1993a, who have extensively analysed the impacts of the choice of ϵ on DEA results). This means that in practical applications it is better to use the two-stage model to identify both the optimal radial factors and slack values, rather than the single stage model (see Chang and Guh, 1991, that also draw on the problems around the choice of ϵ).

The second stage model in (2.16) corresponds to an important model in the DEA literature. If on the right hand side of (2.16) observed values are used instead of target values, the resulting model is the additive model as first introduced by Charnes et al. (1985b). Though the additive model does not provide a final efficiency measure, it is important in two respects, (i) it allows the identification of Pareto-efficient units – those which have a zero sum of slacks in the optimal solution of the model, and (ii) it is one of the first non-oriented models introduced in the literature. The Farrell input and output models presented previously are radial and oriented: they sought either input reduction or output expansion but not both. This imposes very strong assumptions at the beginning of any efficiency measurement: the assumption that only inputs or only outputs are controllable by the decision maker and improvement shall be sought only on these controllable factors. In reality, however, it may happen that at least some inputs and outputs are controllable requiring, therefore, measures that seek improvements on both sides (i.e. non-oriented measures). The additive model is such a measure. Other non-oriented measures will be referred in more detail in subsequent chapters of this study.

2.3.1 Economic Measures of Efficiency

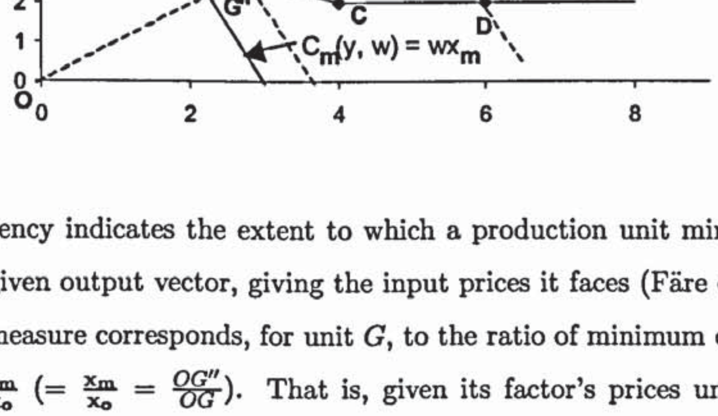
The efficiency measures defined previously do not depend on prices. Indeed, input and output vectors are usually specified in physical units in production models. Efficiency models that use economic variables such as prices, costs, revenues or profits are known as economic models (Färe and Primont, 1995). Two important functions that can be specified are the cost and revenue functions. The first assumes a cost minimising behaviour of the production unit and the second assumes a revenue maximising behaviour of the production unit.

Lets assume that $\mathbf{w} \in R^m_+$ is the input price vector, and $\mathbf{p} \in R^s_+$ is the output price vector. The cost function is defined by $C(\mathbf{y}, \mathbf{w}) = \mathbf{wx} = \min\{\mathbf{wx} \mid \mathbf{x} \in L(\mathbf{y})\}$, corresponding to the minimum expenditure required to produce output vector $\mathbf{y}$ at input prices $\mathbf{w}$ (e.g. Färe et al., 1994a).

The revenue function is given by $R(\mathbf{x}, \mathbf{p}) = \mathbf{py} = \max\{\mathbf{py} \mid \mathbf{y} \in P(\mathbf{x})\}$, representing the maximum revenue that can be generated from input vector $\mathbf{x}$ at output prices $\mathbf{p}$.

With price information it is possible to compute, aside from technical, other measures of efficiency. In his seminal work Farrell (1957) referred to overall (cost or revenue) and allocative efficiency measures. Figure 2.8 illustrates the meaning of such measures. A cost

Figure 2.8: Economic Efficiency Measurement



measure of efficiency indicates the extent to which a production unit minimises the cost of producing a given output vector, giving the input prices it faces (Färe et al., 1985). In Figure 2.8 this measure corresponds, for unit G, to the ratio of minimum cost to observed cost: $\frac{C_m}{C_o} = \frac{\mathbf{wx_m}}{\mathbf{wx_o}}$ $(= \frac{\mathbf{x_m}}{\mathbf{x_o}} = \frac{OG''}{OG})$. That is, given its factor's prices unit G should be producing at point B (or A) and not at its current levels. Point B is overall efficient because it lies simultaneously on the technological frontier and on the minimum cost line.

As seen previously, technical efficiency of unit G is given by $\frac{OG'}{OG}$. Note that this value is equivalent to the ratio between the cost at the technical efficient point and observed cost ($\frac{C_T}{C_o} = \frac{\mathbf{w}\mathbf{x}_T}{\mathbf{w}\mathbf{x}_o} = \frac{\mathbf{x}_T}{\mathbf{x}_o} = \frac{OG'}{OG}$). For example, a technical efficiency value of 50% means at the same time that when inputs decrease by 50%, so that the technical efficient frontier is reached, costs also reduce by 50%. Assume that unit G has eliminated its technical inefficiency by moving to point G'. This point is not cost efficient when compared to point B that has a lower cost. A movement from point G' to point B implies an adjustment in the mix of inputs of unit G that further reduces its costs. Note that by keeping the same mix unit G' could not reduce further its costs without getting out of the technological set. The adjustment from point G' to the point of minimum cost, B, represents allocative efficiency. The allocative efficiency is also called price efficiency, as it measures the extent to which a technically efficient point such as G' falls short of achieving minimal cost because it fails to make the substitutions (or reallocations) involved in moving from G' to B (Cooper et al., 2000). This measure can also be expressed in terms of a ratio between the minimum cost at point B and the cost at the technical efficient point G' ($\frac{C_m}{C_T} = \frac{\mathbf{w}\mathbf{x}_m}{\mathbf{w}\mathbf{x}_T}$). Note that this cost ratio is equivalent to the quantity ratio $\frac{OG''}{OG'}$. We can now establish the following relationship:

$$\text{Cost efficiency } (\tfrac{OG}{OG''}) = \text{ technical efficiency } (\tfrac{OG}{OG'}) \times \text{ allocative efficiency } (\tfrac{OG'}{OG''})$$

It is important to stress that cost efficiency does not have an immediate interpretability in terms of input reductions, although the inverse is always true, i.e., radial input reductions always have a cost interpretability (e.g. Kerstens and Vanden-Eeckaut, 1995). In fact, if we multiply all the inputs of G by the overall cost efficiency measure the resulting input targets would be located exactly at G'', which is a point outside the efficient boundary and thus non-attainable. This point is therefore used only as a reference to calculate cost efficiency and should not be interpreted as the minimum cost target for unit G. The minimum cost target for unit G corresponds to point B.

The computation of cost efficiency involves solving the minimum cost model (2.18), where the decision variables are λ_j and x_i.

$$
\begin{aligned}
&\min\Big\{C_m = \sum_{i=1}^{m} w_{io}x_i \mid \\
&\sum_{j=1}^{n}\lambda_j\, y_{rj} - s_r = y_{ro} \qquad r = 1,\dots,s \\
&\sum_{j=1}^{n}\lambda_j\, x_{ij} + e_i = \; x_i \qquad i = 1,\dots,m \\
&\lambda_j \in S, \quad s_r,\ and\ e_i \geq 0\Big\}
\end{aligned}
\tag{2.18}
$$

The solution of (2.18) results in optimal input levels x_i^* yielding minimum cost C_m^*. Cost efficiency is therefore calculated as C_m^*/C_o, and allocative efficiency can be calculated by decomposition: the ratio of cost efficiency by technical efficiency.

The revenue efficiency model is shown in (2.19).

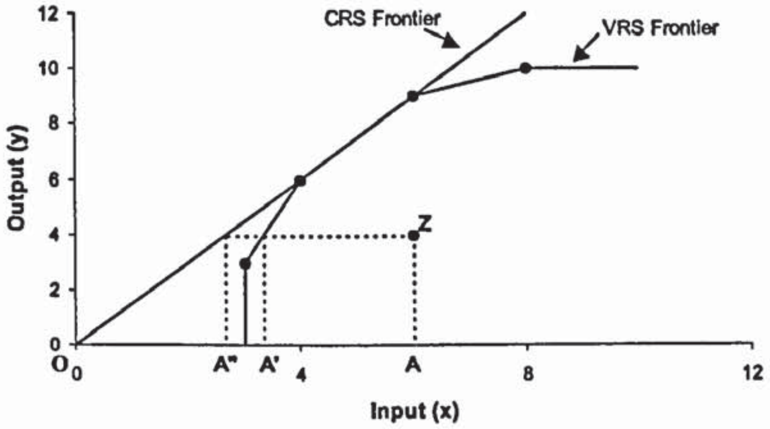
$$
\begin{aligned}
&\max\Big\{R_M = \sum_{r=1}^{s} p_{ro}y_r \mid \\
&\sum_{j=1}^{n}\lambda_j\, y_{rj} - s_r = y_r \qquad r = 1,\dots,s \\
&\sum_{j=1}^{n}\lambda_j\, x_{ij} + e_i = \; x_{io} \qquad i = 1,\dots,m \\
&\lambda_j \in S, \quad s_r,\ and\ e_i \geq 0\Big\}
\end{aligned}
\tag{2.19}
$$

The solution of (2.19) results in optimal output levels y_r^* yielding maximum revenue R_M^*. Revenue efficiency is therefore calculated as $\frac{R_o}{R_M^*}$. Allocative efficiency can be calculated as the ratio of revenue efficiency and technical efficiency, where the latter can be calculated through model (2.12).

2.3.2 Scale Efficiency and Returns to Scale in DEA

We mentioned earlier that a technological set can exhibit different characteristics in terms of returns to scale. CRS and VRS are two of the most used RTS assumptions. A comparison between efficiency measures obtained under each of these technologies allows one to draw conclusions about scale efficiency and returns to scale (Banker et al., 1984). Figure 2.9 illustrates, for a single input/output case, the CRS and VRS boundaries (these are also called CCR and BCC boundaries, respectively, in respect for their developers, Charnes Cooper and Rhodes, and Banker, Charnes and Cooper, respectively). Unit Z is

Figure 2.9: Technical and Scale Efficiency Measures



inefficient both under CRS and VRS. The input measure of technical efficiency calculated in relation to the VRS frontier is $E_{VRS} = OA'/OA$, while the measure of technical efficiency calculated in relation to the CRS frontier is $E_{CRS} = OA''/OA$. For the single input/output case CRS efficient units are those presenting maximum productivity, measured by the ratio of output to input. Such units are denoted most productive scale size[6] (mpss) by Banker (1984). An efficiency measure calculated in relation to a VRS technology compares production units with efficient units (or a composite of efficient units) presenting a similar scale. This does not happen in CRS technologies where units are compared with the optimal scale size. As a result, the distance between the VRS and the CRS frontiers represents scale efficiency, i.e. the shortfall of the VRS frontier to optimal scale size at a given point. The scale efficiency of unit Z is, therefore, given by OA''/OA', which equals E_{CRS}/E_{VRS}.

The global measure of efficiency (measured in relation to the CRS frontier) is, therefore, a composite of pure technical efficiency (measured in relation to the VRS frontier) and scale efficiency (SE), as $\frac{OA''}{OA} = \frac{OA'}{OA} \times \frac{OA''}{OA'} \Leftrightarrow E_{CRS} = E_{VRS} \times SE$.

If output oriented efficiency measures were used in the illustrative example in Figure 2.9 the values of scale efficiency for each unit would be different. CRS efficiency is the same irrespective of the model orientation, but VRS efficiency is not. This means that scale efficiency is always dependent on the model orientation.

[6]According to Banker (1984, pp. 39-40) the CRS technical efficiency of a given DMU is equal to 1 if and only if that unit is a mpss. In addition, a CRS technical efficiency measure of 1 implies that we are in the presence of a mpss. Mpss units are therefore scale efficient units lying on a region of constant returns to scale. In addition, a vector $(\mathbf{x}, \mathbf{y}) \in T$ is a mpss if and only if for all production possibilities $(\beta\mathbf{x}, \alpha\mathbf{y}) \in T$ we have $\frac{\alpha}{\beta} \leq 1$.

Apart from scale efficiency, qualitative information regarding the type of RTS that apply at specific parts of the efficient frontier can also be obtained in DEA models. A variety of methods has been proposed to characterise the type of returns that apply on the efficient frontier. Some of these are based on solutions from envelopment models and others are based on solutions from multiplier models. Banker et al. (1984) proposed the use of the multiplier VRS models in (2.13) and (2.14) to characterise RTS. Based on the optimal value of u_o, represented by u_o^*, Banker et al. (1984) presented a classification of the type of RTS. This initial characterisation did not account for the possibility of multiple optimal solutions. This drawback was overcome in Banker and Thrall (1992), who provided the following characterisation of returns to scale according to the value of u_o^* (see also Cooper et al., 2000):

- $u_o^* < 0$ for all optimal solutions $\Rightarrow$ IRS;
- $u_o^* = 0$ in any optimal solution $\Rightarrow$ CRS;
- $u_o^* > 0$ for all optimal solutions $\Rightarrow$ DRS.

Banker and Thrall (1992) developed some auxiliary linear programs that enabled them to deal with multiple optimal solutions through the calculation of intervals for the variable u_o.

It is also possible to estimate returns to scale based on envelopment models defined in relation to a CRS technology. In this case we need to look at the sum of the optimal lambda values. Banker and Thrall (1992) proposed the following rules for classifying the type of returns to scale prevailing at a given boundary point.

- $\sum_{j=1}^{n} \lambda_j^* < 1$ for all alternate optima $\Rightarrow$ IRS;
- $\sum_{j=1}^{n} \lambda_j^* = 1$ in any alternate optima $\Rightarrow$ CRS;
- $\sum_{j=1}^{n} \lambda_j^* > 1$ for all alternate optima $\Rightarrow$ DRS.

The procedure for testing whether the above conditions are satisfied in all alternate optima consists in a second step linear programming problem (e.g. Cooper et al., 2000; Banker et al., 1996a).

Another method that can be applied to characterise RTS is that of Färe et al. (1985), which consists in obtaining three efficiency estimates in relation to three technological RTS specifications: CRS, VRS, and NIRS. From the efficiency measures obtained from each of these models, conclusions can be reached concerning returns to scale:

- If CRS, VRS and NIRS models yield exactly the same efficiency measure, then the unit lies, or is projected, on a boundary region exhibiting CRS;
- If CRS and NIRS efficiency measures are equal and lower than the VRS efficiency measure, then the unit lies, or is projected, on an IRS region of the boundary;
- If VRS and NIRS efficiency measures are equal and higher than the CRS efficiency measure, then the unit lies, or is projected, on a DRS region of the boundary;

The Färe et al. (1985) method has the advantage of being unaffected by the existence of multiple optimal solutions. Its main disadvantage seems to be the need to solving three DEA problems (Seiford and Zhu, 1999).

The methods mentioned hitherto provide well defined RTS classifications only for production units lying on the efficient frontier. For inefficient units "productivity changes due to returns to scale are confounded with productivity changes due to inefficiency elimination" (Banker and Thrall, 1992, p. 82). In addition the RTS classifications of inefficient units depends on their projection on the efficient boundary. This might yield different results depending on the orientation of the model (towards input contraction or output expansion). Seing this fact as a limitation, some authors proposed other approaches to characterise RTS. For example, Golany and Yu (1997) developed a model which may result in a strict characterisation of the RTS that apply at a given point of the frontier, but it may also happen that no estimate of RTS for a particular unit can be obtained. This last situation happens when the RTS results accruing from input and output oriented models are contradictory (this clearly only happens for inefficient units). The authors also analysed the situation of extreme efficient units whose returns to scale situation might be ambiguous. A model similar to that in Golany and Yu (1997) is presented in Cooper et al. (1996) who propose the use of the most productive scale size (mpss) definition to characterise returns to scale. Another approach is that of Fukuyama (2003) who proposed the use of a non-radial efficiency measure to estimate RTS. The chosen efficiency measure was in this case the directional distance model of Chambers et al. (1996a, 1998). The use of such a measure eliminates to a certain extent the ambiguity of RTS classifications resulting from input and output oriented models. Nevertheless, RTS classifications are still dependent on the frontier point where inefficient observations are projected.

Several other references concerning the estimation of RTS can be found in the literature. For example Zhu and Shen (1995), introduce a simple approach that eliminates the need to examining all alternate optima, Färe and Grosskopf (1994b) contrast their method to the Banker and Thrall's, and Banker et al. (1996a) summarise some methods and prove their equivalence to the method of Färe et al. (1985). Appa and Yue (1999) proposed an interesting procedure to determine unique scale efficient targets based on the concept of mpss. Zhu (2000) extended this approach calculating scale efficient targets that correspond to either the largest or the smallest mpss. Two interesting literature reviews on returns to scale are presented by Seiford and Zhu (1999) and Löthgren and Tambour (1996).

An issue closely related to RTS is that of scale elasticity, that is, the proportionate increase in outputs resulting from proportionate increase in inputs (see Banker and Thrall, 1992; Fukuyama, 2000). Banker and Thrall (1992) propose the calculation of scale elasticity $\bar{\varepsilon}$ through the lower and upper bounds of u_o^*, where $\frac{1}{1-u_o^-} \leq \bar{\varepsilon} \leq \frac{1}{1-u_o^+}$. Førsund and Hjalmarsson (1979) and Førsund (1996) also provide an analysis of scale elasticity and put forward its relationship with scale efficiency. For more details on the calculation of scale elasticity in DEA see also Cooper et al. (1996) and Førsund and Hernaes (1994).

2.3.3 Non-Discretionary Factors

Basic DEA models referred to in the previous sections assume that either all inputs or outputs (depending on the model orientation) are discretionary or controllable factors. However, in most real situations there are non-discretionary or exogenously fixed factors that managers cannot control. Such factors are, in most cases, environmental characteristics of production units that affect their performance but cannot be controlled by them.

Fried et al. (1999) introduced a classification of methods that are used to deal with environmental or non-discretionary variables: (i) the frontier separation approach, (ii) the all-in-one approach, and (iii) the two-stage approach. We use the same classification here to briefly review the various methodologies that have been used in the literature to deal with non-discretionary factors (see also Coelli et al., 1999).

The Frontier Separation Approach

The frontier separation approach groups production units according to some criteria (usually categorical) and performs separate efficiency assessments for each of these groups. An assessment of the pooled groups is then also undertaken in order to ascertain the impact of the criteria on efficiency. A first application of this approach can be seen in Charnes et al. (1981) where schools running under the follow through program were compared to those which were not run under this program. Another example applied to schools can be seen in Portela and Thanassoulis (2001) where two levels of aggregation were considered in assessing pupils efficiency: the school and the school type.

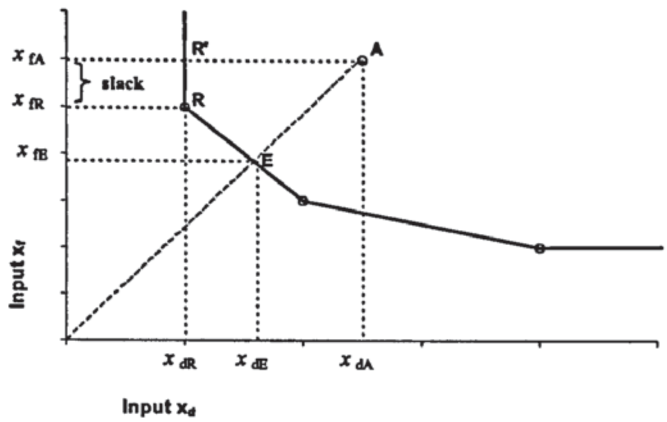
The approach developed by Banker and Morey (1986b) for dealing with categorical variables can also be included under frontier separation methods. This approach consists of introducing a set of dummy variables concerning the categorical factor so that units belonging to a given group can only be compared with units in less favourable groups than the unit being assessed. In this sense we can visualise a set of frontiers that are building up (see also Charnes et al., 1994, pp. 51-54). This procedure can, according to Banker and Morey (1986b), be used both when non-discretionary factors are continuous or categorical variables, as any continuous variable can be transformed in a categorical variable. Some improvements in the Banker and Morey (1986b) approach can be found in Kamakura (1988) and Rousseau and Semple (1993).

There are at least two disadvantages in this procedure that are worth noting. First, it implies the various categories to be ordered hierarchically, which is not always natural (Førsund, 2001). Secondly, when there are several criteria the homogeneous groups of units may be very small. As the discriminant power of DEA depends on the number of units and on the number of input/output variables considered in the assessment, the smaller the group the smaller is the discrimination between production units achieved by DEA. Staat (1999) refers also to the problem associated with differing sample sizes across groups (see also Zhang and Bartels, 1998, on this subject), and to the problem of comparability between the resulting efficiency scores. Indeed, when production units are grouped according to some criteria only efficiency rankings inside the same group can be meaningfully compared.

The All-in-One Approach

The all-in-one approach includes the non-discretionary variables directly into the DEA efficiency assessment (this obviously excludes the possibility of categorical variables to be treated under this approach). The usual procedure is to allow these variables to affect the shape of the production frontier (which necessarily happens when they are considered as inputs or outputs of the efficiency assessment) but they are kept constant in the calculation of the efficiency score. The best known procedure inside this general approach is that developed by Banker and Morey (1986a). The proposed DEA model, in its envelopment form, consists in associating the expansion (contraction) factor only to discretionary outputs (inputs) in a first phase, and by maximising in the second phase only the sum of the slacks associated to the discretionary factors. Non-discretionary factors are left aside in terms of finding the optimal radial contraction or expansion factor and in terms of finding the optimal slack values. Such a treatment of non-discretionary factors naturally results in non-radial efficiency measures, however this aspect has not been stressed in the literature. Taking the example provided by Banker and Morey (1986a) in Figure 2.10 we can see that the radial efficiency measure of unit A (if all inputs were discretionary) would be x_{dE}/x_{dA}. When we consider input x_f as non-discretionary the efficiency measure will be x_{dR}/x_{dA},

Figure 2.10: Banker and Morey Example



which is a non-radial measure. That is, the projection on the efficient boundary does not consider the reduction of the non-discretionary input and as such the discretionary input's reduction is greater than in radial movements. It is noticeable from Figure 2.10 that there is a positive slack associated to the non-discretionary input. The value of such a slack is only informative as the non-discretionary input cannot be reduced by the slack amount so that the efficient subset of the technological frontier is attained. This means that the relevant subset for projection should be such as to consider efficient also those frontier parts where strong disposability of the non-discretionary factor apply.

Golany and Roll (1993) extended the work of Banker and Morey (1986a) by addressing the question of simultaneous non-discretionary inputs and outputs and also partially non-discretionary factors. Cooper et al. (2000) extended the Banker and Morey (1986a) approach to the additive model.

Some criticisms to the Banker and Morey (1986a) model can be found in Ruggiero

(1996) and Ruggiero (1998), where the author advocates that the Banker and Morey's model does not properly restrict the reference set. According to this author the enforcement of convexity to the non-discretionary factors "leads to improper restriction of the production possibility sets and distorted efficiency measurement" (Ruggiero, 1998, p. 463). Another criticism to the Banker and Morey (1986a) model concerns the fact that targets may be constructed from any set of units. That is a unit in a "medium" environment may have as peers units in a "good" environment and units in a "bad" environment. One way of avoiding this is to consider in the reference set only units with environmental conditions that are equal or worst to those of the unit being assessed. This is the approach proposed by Ruggiero (1996) [see also Staat (1999)], which is, however, very close to the approach of Banker and Morey (1986b) meaning that it actually reduces to a frontier separation approach.

Muñiz (2002) also criticises the Banker and Morey (1986a) approach pointing out the fact that the resulting frontier is exactly the same as it would be if the non-discretionary factors were considered controllable. As a result, the environmental conditions have no influence on the efficiency status of production units and only those that are inefficient are actually penalised by the consideration of some factors as being non-discretionary.

Besides these criticisms, the Banker and Morey (1986a) approach to treat non-discretionary factors is still the most often used approach given its simplicity and its inclusion of all factors (discretionary or not) in a single DEA model.

The Two-Stage Approach

A different approach to handle non-discretionary factors was introduced by Ray (1988) and further developed in Ray (1991). According to this author non-discretionary or environmental factors should not be included in a DEA assessment. A DEA assessment should include only controllable factors, and then in a second phase a regression model should be used to account for the uncontrollable factors. The difference between the efficiency score estimated through the regression model and the efficiency score obtained from the DEA analysis is interpreted by Ray (1991, p. 1627) as the extent of managerial inefficiency not caused by external factors. That is, inefficiency is calculated as the shortfall of the DEA efficiency score from the estimated efficiency score and not from 1. As pointed out by Ruggiero (1998) this approach requires the specification of a functional form to the regression model meaning that a mis-specification may distort the results (see also McCarty and

Yaisawarng, 1993).

One problem of the two-stage approach relates with the possible correlation between the input/output factors used to calculate the DEA efficiency scores and the independent variables used in the regression model (see Grosskopf, 1996). Another problem was referred to by Xue and Harker (1999) as the dependency problem. That is, the DEA efficiency scores are dependent on each other, which "violates a basic assumption required by regression analysis: the assumption of independence within the sample" (Xue and Harker, 1999, p. 3).

Other Alternatives

Apart from the above mentioned methods to deal with non-discretionary factors there are other alternatives that have been proposed in the literature. In general these alternatives involve multi-stage procedures. This is for example the case of the approach proposed by Ruggiero (1998) that links the approach of Ray (1991) and Ruggiero (1996) trying to overcome its weaknesses. Indeed, in the Ruggiero (1998) model the non-discretionary factors are considered in a weighted 'index' that assigns different importance to the various factors according to the results of a regression analysis. This model only uses the regression to calculate the weights to assign to the environmental factors and not to calculate an adjusted measure of efficiency. This means, according to the author, that no distributional assumptions are made regarding efficiency.

Another multi-stage model was introduced by Fried et al. (1999), and it deals with environmental factors in four stages. According to Fried et al. (1999) this procedure has the advantage of accounting for all sources of inefficiency as it regresses slacks (through a Tobit model) on environmental factors instead of radial efficiency scores. In addition, it also has the advantage of not requiring a pre-specification of environmental factors as inputs or outputs (which is required, for example, in an all-in-one approach).

2.3.4 Super-Efficiency

DEA defines an efficient frontier constituted by the best performer production units. Such best performers may in some cases be outliers whose distance from the bulk of units under analysis is considerable. "DEA is in many ways, an outlier-based method. DMUs which are outliers in terms of low input relative to output levels map out the efficiency boundary" (Thanassoulis, 1993, p. 1132). This means that special attention should be devoted to outliers since some of them may be defining an efficient frontier that is not reasonably achievable by most of the units under analysis.

Outliers, also known as influential units, are observations that present a unique combination of variables and might, for that reason, distort the analysis. In such cases, the exclusion of influential units should, in principle, be undertaken.

A number of methods have been proposed to detect influential observations. Most of these methods apply the super-efficiency concept first introduced by Andersen and Petersen (1993). The super-efficiency model establishes an efficient frontier where the unit being assessed cannot be in. This may result in efficiency scores greater than one for those units lying originally on the efficient frontier. Such efficiency scores allow discrimination and ranking of originally 100% efficient units. Mathematically the changes required in basic DEA models to exclude the unit being assessed from the efficient frontier are quite straightforward. Consider the set J_o that includes all the units under analysis except unit o being assessed. In an envelopment model the terms $\sum_{j=1}^{n} \lambda_j y_{rj}$ and $\sum_{j=1}^{n} \lambda_j x_{ij}$ are such that $j \in J_o$ (see models (2.11) and (2.12)). In the dual multiplier model this corresponds to eliminating the constraint $\sum_{i=1}^{m} v_i x_{ij} - \sum_{r=1}^{s} u_r y_{rj} + u_o \leq$ (or $\geq$) 0 for $j = o$ (see models (2.13) and (2.14)).

Dusansky and Wilson (1994, 1995) checked for influential observations in the data by deleting efficient units one at a time and recomputing the efficiency scores. The efficiency scores resulting after each deletion were averaged and compared with the average efficiency scores of the full sample (see also Wilson, 1995). "Sums of absolute and squared distances in efficiencies across the DMUs were also compared" in order to check for outliers. The main problem with this approach is that more than one observation can be influential and a set of influential observations may be determining the shape of the frontier. If one of these is removed only the shape of the frontier will change but not its relative position from the other units. Recognising this problem, Dusansky and Wilson (1994) (see also Dusansky and Wilson, 1995) also excluded pairs of observations, after which the same type of calculations were undertaken. The criteria to consider under which circumstances an observation was indeed an outlier were rather subjective. Indeed, there are no statistical methods to indicate when changes in average efficiency are statistically significant, especially when the rankings are not independent, as it is the case in DEA.

Another approach to exclude influential units was proposed by Thanassoulis (1999b), applied to the school context where the production units considered were pupils. Super-efficient pupils were detected through the use of the super-efficiency model of Andersen and Petersen (1993). The decision on exclusion of super-efficient pupils was based on an iterative process where some thresholds were specified so that the resulting efficient frontier can be said to reflect what pupils can really achieve.

A different methodology was used by Färe et al. (1989), which has been dubbed jackknifing. This methodology consists of dropping one unit at a time in the calculation of efficiency. This results for each unit o in a set of $n-1$ efficiency scores calculated when each unit $j \neq o$ is eliminated from the assessment. For these $n-1$ efficiency scores some descriptive statistics can be computed. According to Färe et al. (1989, p. 414) "this technique allows us to employ statistical tests and reduces the effects of outliers in the results". Accordingly, small standard deviations are understood as suggesting the absence of outliers in the sample. The jackknifing technique is criticised by Ondrich and Ruggiero (2002), where the author argues that jackknifing is not useful in the detection of outliers and it cannot deal with multiple outliers. In fact for efficient observations the jackniffing technique results in $n-1$ efficiency scores of 1 and therefore the resulting standard deviation is zero. As outliers are particularly meaningful when they lie on the estimated production frontier Ondrich and Ruggiero (2002) conclude that this technique cannot correctly identify outliers.

In summary a combination of the above methods can be applied to detect outliers or influential observations. No method is perfect and ultimately all of them are subjective. What is important to keep in mind is that in some circumstances the frontier is ill-defined because of influential observations and their exclusion from the analysis is important to guarantee a fair efficiency assessment.

2.3.5 Weights Restrictions

DEA gives full flexibility to production units in choosing the weights assigned to inputs and outputs. Such flexibility can be regarded as an advantage: "clearly if you are free to employ your own (optimal) weights and somebody else uses them to beat you at your own game then a strong statement is being made" (Tofallis, 1996, p. 362). However, it can also be regarded as a disadvantage especially as long as efficient units are concerned. Indeed, the weight's freedom of choice may result in production units completely neglecting certain inputs and/or outputs and by doing so becoming efficient. This means that the final ratio of efficiency might be meaningless, especially if a number of factors is neglected.

Early practical applications of DEA called attention to the fact that efficient units should be treated with caution because in some cases the weights attributed to input and output factors were suspicious. For example, Thanassoulis et al. (1987) called attention to the fact that some efficient local tax offices placed most of the weight on some outputs while completely neglecting others.

Additional restrictions to the multiplier DEA model can be imposed to avoid the allocation of zero weights to some factor's of production. Such weights restrictions (WRs) incorporate in principle prior views of decision makers (DMs) concerning the worth of each factor of production. WRs may also be used with other purposes such as improving the discrimination between efficient units, or to reduce the variability of weights across DMUs (see Allen et al., 1997, for details).

Absolute Weights Restrictions

The most immediate form of placing restrictions on the weights is to restrict them to be lower and/or higher than a given value. These restrictions can assume the form expressed in (2.20), where v_i stands for input weight and u_r stands for output weight. Such constraints are called absolute weight restrictions.

$$b_i \leq v_i \leq a_i \qquad c_r \leq u_r \leq d_r \tag{2.20}$$

The first use of this type of weights can be seen in Dyson and Thanassoulis (1988) in an application to local tax offices. Cook et al. (1994) also used direct restrictions on the weights in an application to highway maintenance patrols, and Roll et al. (1991) explored in some detail the implications of imposing absolute WRs to DEA models. Absolute WRs have some problems, one of which is the potential infeasibility of DEA models with these WRs. Podinovski (2001, p. 575) refers and illustrates this fact, also claiming that restricted models "may not identify the maximum relative efficiency of the assessed DMU correctly". In a sense, this means that, the DEA weights under absolute WRs may not enable a DMU to appear in the best possible light relative to other DMUs (see also Podinovski and Athanassopoulos (1998) and Podinovski (1999)). Another problem with absolute WRs relates with the interdependence between weights, i.e. "an upper bound on one weight imposes lower bounds on the other weights" (Roll et al., 1991, p. 5). At the same time the interpretation of the bounds placed on the weights under absolute WRs is not easy, since weights are especially meaningful on a relative basis[7]. Some procedures using virtual input and output weights[8] were developed to overcome the latter problem, such as the approach of Wong and Beasley (1990). Virtual input/output weights have the advantage of being invariant to units of measurement, which does not happen with the weights of the multiplier model. This means that care is needed when analysing these weights because a higher or lower weight does not necessarily mean a higher or lower importance attached to a specific factor, since all depends on the units of measurement of the factors. Restrictions on virtual weights have, therefore, the advantage of being units invariant, but they usually require more complex and computationally expensive procedures, since virtual weights are DMU specific (e.g. Allen et al., 1997; Thanassoulis, 2001).

Relative Weights Restrictions

The most common relative WRs are the Assurance Regions (ARs), first developed by Thompson et al. (1986). ARs can assume the form in (2.21) or (2.22), where one of the input or output weight was chosen as numeraire (say v_1 and u_1).

$$
\begin{aligned}
b_i v_1 \;\leq\; v_i \;\leq\; a_i v_1 \qquad & i = 1, \ldots, m \qquad \text{(input cone)} \\
c_r u_1 \;\leq\; u_r \;\leq\; d_r u_1 \qquad & r = 1, \ldots, s \qquad \text{(output cone)} \qquad (2.21) \\
v_i \;\geq\; e_i u_r \qquad & \qquad\qquad\qquad\; \text{(linked cone)} \qquad (2.22)
\end{aligned}
$$

The set of constraints in (2.21) are called ARI and constitute separable input and output cones. The constraints in (2.22) form AR-II, which link input and output multipliers. The specification of bounds in ARs (like the scalars a_i, b_i, c_r, d_r, and e_i) may be based on input and output prices when these are known with precision (e.g. Cooper et al., 2000). However, the main advantage of using price information in ARs arises when prices are not precisely known, and price intervals are used in the construction of lower and upper AR bounds. Thompson et al. (1990) argue that the inclusion of price information, even imprecise goes beyond technical efficiency, representing a step towards cost or revenue efficiency measurement (see also Thanassoulis et al., 2003).

The imposition of ARs on the multiplier model implies the addition of new variables

[7]Indeed, ratios os weights are equal to marginal rates of transformation (substitution) between outputs (inputs).

[8]Virtual input or output weights are the product of optimal weights, resulting from the multiplier model, and observed level of inputs or outputs.

in the envelopment model. Such variables can be seen as new slack variables or new units added to the data set (e.g. Roll et al., 1991; Thanassoulis and Allen, 1998). Alternatively, as mentioned in Thanassoulis et al. (2003), ARs may also be interpreted as new facets that are added to the original PPS, therefore changing the efficient frontier against which efficiency is measured.

Another type of relative weight restrictions is the Cone Ratio (CR) procedure introduced by Charnes et al. (1990). The CR approach consists of transforming the input/output vector from $(\mathbf{x}, \mathbf{y})$ to $(\mathbf{x}', \mathbf{y}') = (A\mathbf{x}, B\mathbf{y})$, such that the standard multiplier or envelopment models can be used on the transformed data set. The matrixes A and B can be specified in a number of ways, namely in a way that assures the equivalence between the CRs and ARs (see for example Charnes et al., 1990). In Charnes et al. (1990) and Brockett et al. (1997a) the matrixes A and B were specified based on the multipliers of role model or excellent production units (banks in that case). That is, the weights of excellent banks were first found through an unrestricted DEA model and then these weights were used to assess the efficiency of all the other banks under analysis. Note that ARs are a special case of the more general CR procedure.

Addition of Unobserved Production Units

As mentioned above, changes in the efficient frontier resulting from the imposition of ARs may be interpreted as if new production units were added to the original data set (Roll et al., 1991). This interpretation gave raise to another strand of literature that introduces unobserved units to the original data set. The first study to enlarge the reference set was that of Golany and Roll (1994), where standards were introduced in the DEA assessments. The main difficulty with this approach relates with the establishment of standards. The authors refer to this problem, but no guidelines on how these standards are actually to be generated were provided.

Allen and Thanassoulis (1996) and Thanassoulis and Allen (1998) developed an approach that introduces Unobserved Decision Making Units (UDMUs) into the reference set. Such approach was proved to be equivalent to the addition of weight restrictions to the multiplier model. The most difficult aspect of the UDMUs approach is the actual creation of the unobserved production units. This is done in Thanassoulis and Allen (1998) through the identification of a subset of units, called 'anchor', whose inputs/output levels are changed so that the resulting boundary better envelops all data points[9]. The UDMUs are constructed from anchor DMUs and reflect the DMs value judgments on the trade-offs between inputs and/or between outputs.

The enlarged data set (with observed and unobserved units) is seen in the UDMUs approach as the final data set against which efficiency is measured. This can be done through standard DEA models and standard procedures for identifying targets are also valid. This means that radial targets may lie outside the original PPS but they are deemed feasible as long as they represent trade-offs that accord with the DM's value judgments.

The imposition of weights restrictions of any of the types mentioned above raises some interpretation issues as to the efficiency scores and target levels obtained from restricted DEA models. We will not detail on these issues here for the sake of brevity. In Allen et al. (1997) and Thanassoulis et al. (2003) these issues are dealt with in some detail.

2.3.6 Efficiency Over Time

In the presence of time series data one can analyse the evolution of efficiency over time. This can be done by plotting efficiency values in each time period and analysing how efficiency has evolved over time. Such an analysis is, however, incomplete since not only efficiency can change from period to period, but it is also likely that the efficient frontier, against which efficiency is measured, changes from period to period. For this reason, it is important to compare efficiency measures when these are calculated in relation to the same time frontier. In addition, it is also of interest to analyse the changes in the technological frontier from period to period.

There are some methods that can be used to perform a temporal analysis in DEA. The one we will focus here with some detail is the Malmquist productivity index. Another method that can be pointed out is the DEA window analysis first introduced by Charnes et al. (1985a). It works by defining a window (set of time periods whose observations are gathered together), and assessing efficiency for units in each window. Then a new window is created with the first time period observations deleted and the subsequent observations added. The process is continued until no more windows can be created. By comparing efficiency scores of the same unit in different periods of time one may infer about the way its efficiency is evolving.

[9]Anchor DMUs are those efficient DMUs "which delineate the DEA-efficient from the DEA-inefficient parts of the PPS boundary" (Allen and Thanassoulis, 1996, p. 9).

Another approach to analyse efficiency evolution is by Tulkens and Vanden-Eeckaut (1995) who define three production reference sets: (i) 'contemporaneous', which includes only units in the same time period, (ii) 'sequential', which includes all the units up to the time period being analysed, and (iii) 'intertemporal', which includes units from all time periods under analysis. The number of production frontiers defined is different in each case: as many as the number of time periods in the contemporaneous and sequential analysis, and a single frontier in the intertemporal analysis. Tulkens and Vanden-Eeckaut (1995) also propose ways of measuring progress and regress for each observation as a way to analyse the changes in the production frontiers over time. This is in general done through models that assess a given unit in time t in relation to a production frontier that relates to another time period. The procedure of Tulkens and Vanden-Eeckaut (1995) is closely related to Malmquist productivity indexes, that will be detailed next.

Malmquist Indexes

The Malmquist productivity index was introduced by Caves et al. (1982), and first used in the DEA literature by Berg et al. (1992) and Färe et al. (1994b). A Malmquist index is in fact a ratio of productivity ratios observed in different periods of time. If we compute a productivity ratio $P_t = \frac{y_t}{x_t}$ in period t, then the ratio $\frac{P_{t+1}}{P_t}$ measuring the change in productivity from t to $t+1$, is a Malmquist productivity index. If this ratio is greater than 1 productivity increased, and if it is lower than 1 productivity decreased.

When a number of factors are being considered in the analysis the calculation of total factor productivity indexes is not easy because, in the absence of prices, there is no meaningful way to aggregate different input and/or output values. A means to solve this problem is to use radial distance functions like those mentioned earlier in this Chapter. Given an efficiency score $\theta^t_{j_t}$ indicating the radial efficiency of unit j as observed in period t and assessed in relation to the technology of period t (superscript), a Malmquist productivity index can be computed as $M^t_j = \frac{\theta^t_{j_{t+1}}}{\theta^t_{j_t}}$. In this case the technology against which both observations are being measured is t. Obviously the reference technology can also be $t+1$, which results in $M^{t+1}_j = \frac{\theta^{t+1}_{j_{t+1}}}{\theta^{t+1}_{j_t}}$. The values of these two Malmquist indexes may differ and as such Färe et al. (1994b) consider the geometric mean of both as the Malmquist total factor productivity (TFP) index, that is, $M_j = (\frac{\theta^t_{j_{t+1}}}{\theta^t_{j_t}} \times \frac{\theta^{t+1}_{j_{t+1}}}{\theta^{t+1}_{j_t}})^{(1/2)}$. This ratio can be decomposed into two components measuring efficiency change (or catch up component) and technical change (or boundary shift component) respectively. Efficiency change (EFCH), is simply the ratio of the efficiency of the same unit when evaluated in two different time periods, that is: $\frac{\theta_{j_{t+1}}^{t+1}}{\theta_{j_t}^{t}}$. Technical change (THCH) is given by $\left(\frac{\theta_{j_{t+1}}^{t}}{\theta_{j_{t+1}}^{t+1}} \times \frac{\theta_{j_t}^{t}}{\theta_{j_t}^{t+1}}\right)^{(1/2)}$.

The use of the geometric average for defining a Malmquist TFP index is not necessary as pointed out by Berg et al. (1992) (see also Førsund, 1993). These authors propose an alternative approach where the Malmquist index is calculated in relation to a base period. The base period Malmquist index is a 'chain version' of the Malmquist index, which satisfies the circularity property. Circularity means that "the index from 1 to 3 is equal to the product of the index from 1 to 2 and the index from 2 to 3" (Berg et al., 1992, pp. S215-S216). Malmquist indexes as defined by Färe et al. (1994b) are not circular.

Färe et al. (1994b, p. 74) refer that "in principle, one may calculate Malmquist productivity indexes relative to any type of technology (i.e., satisfying any type of returns to scale)". This means that although the Malmquist productivity index defined before uses generally an efficiency measure, θ, calculated in relation to a CRS technology, one may assume that the technology exhibits other types of returns to scale. In this situation we can further decompose the efficiency change (EFCH) into a pure technical efficiency change (PTECH) and a scale efficiency change (SECH) components, i.e. $\frac{\theta_{j_{t+1}}^{t+1}}{\theta_{j_t}^{t}} = \frac{\beta_{j_{t+1}}^{t+1}}{\beta_{j_t}^{t}} \times \frac{\theta_{j_{t+1}}^{t+1}/\beta_{j_{t+1}}^{t+1}}{\theta_{j_t}^{t}/\beta_{j_t}^{t}}$, where β are VRS efficiency scores, and the scale change is given by the ratio of efficiency scores calculated under CRS and VRS.

The Malmquist decomposition proposed by Färe et al. (1994b) will be called FGNZ and is resumed in (2.23)[10].

$$M_j = \left(\frac{\theta_{j_{t+1}}^{t}}{\theta_{j_t}^{t}} \times \frac{\theta_{j_{t+1}}^{t+1}}{\theta_{j_t}^{t+1}}\right)^{(1/2)} = \frac{\beta_{j_{t+1}}^{t+1}}{\beta_{j_t}^{t}} \times \frac{\theta_{j_{t+1}}^{t+1}/\beta_{j_{t+1}}^{t+1}}{\theta_{j_t}^{t}/\beta_{j_t}^{t}} \times \left(\frac{\theta_{j_{t+1}}^{t}}{\theta_{j_{t+1}}^{t+1}} \times \frac{\theta_{j_t}^{t}}{\theta_{j_t}^{t+1}}\right)^{(1/2)}$$
$$M_j = PTECH \times SECH \times THCH = EFCH \times THCH \tag{2.23}$$

Although the FGNZ decomposition is widely used, it is not free of controversy. For example Ray and Desli (1997) argue that when the technology is not characterised by CRS, then technical change should not be the result of comparing CRS technologies (as above) but the result of comparing VRS technologies. Following this reasoning Ray and Desli (1997) come out with a different decomposition (the RD decomposition) which is arguably more appropriate for VRS technologies. The decomposition of Ray and Desli

[10]There are other possible decompositions to the FGNZ Malmquist index, namely those put forward by Färe et al. (1998) where technical change is decomposed into an output biased technical change component, an input biased technical change component, and a magnitude component. These decompositions will not be detailed here.

(1997) is shown in (2.24), where $S^t_{j_t} = \frac{\theta^t_{j_t}}{\beta^t_{j_t}}$, $S^t_{j_t+1} = \frac{\theta^t_{j_t+1}}{\beta^t_{j_t+1}}$, $S^{t+1}_{j_t+1} = \frac{\theta^{t+1}_{j_t+1}}{\beta^{t+1}_{j_t+1}}$, and $S^{t+1}_{j_t} = \frac{\theta^{t+1}_{j_t}}{\beta^{t+1}_{j_t}}$.

$$M_j = \left(\frac{\theta^t_{j_t+1}}{\theta^t_{j_t}} \times \frac{\theta^{t+1}_{j_t+1}}{\theta^{t+1}_{j_t}}\right)^{(1/2)} = \frac{\beta^{t+1}_{j_t+1}}{\beta^t_{j_t}} \times \left(\frac{S^t_{j_t+1}}{S^t_{j_t}} \times \frac{S^{t+1}_{j_t+1}}{S^{t+1}_{j_t}}\right)^{(1/2)} \times \left(\frac{\beta^t_{j_t+1}}{\beta^{t+1}_{j_t+1}} \times \frac{\beta^t_{j_t}}{\beta^{t+1}_{j_t}}\right)^{(1/2)}$$

$$M_j = PTECH \times SECH \times THCH \tag{2.24}$$

The first term in (2.24) is the pure technical efficiency change, the second is the scale change component, and the third is the technical change or boundary shift effect. Note that differences in relation to the FGNZ approach happen for the technical change effect, which is defined in terms of the VRS technology, and for the scale effect, which is given by a geometric mean of two effects rather than by a simple ratio as in the FGNZ approach.

Färe et al. (1997b, p. 1041) criticised the RD approach based on the argument that the RD technical change component measures the shift in the VRS technology "but that shift is not the change in maximal average product", which is only measured when CRS technologies are assumed. Färe et al. (1997b) also point out the possibility of infeasible DEA models when VRS efficiency is calculated in relation to a different time period technology. Infeasible models lead to computational problems in the RD decomposition. Such does not happen when the FGNZ decomposition is performed because, under CRS, efficiency can always be computed both in relation to the own time period technology and in relation to another time period technology.

There is also another decomposition of interest in the literature that is a kind of mixture between the RD and FGNZ approaches. Such a decomposition was proposed by Wheelock and Wilson (1999) and is shown in (2.25).

$$M_j = \frac{\beta^{t+1}_{j_t+1}}{\beta^t_{j_t}} \times \left(\frac{S^{t+1}_{j_t+1}}{S^t_{j_t}}\right) \times \left(\frac{\beta^t_{j_t+1}}{\beta^{t+1}_{j_t+1}} \times \frac{\beta^t_{j_t}}{\beta^{t+1}_{j_t}}\right)^{1/2} \times \left(\frac{S^t_{j_t+1}}{S^{t+1}_{j_t+1}} \times \frac{S^t_{j_t}}{S^{t+1}_{j_t}}\right)^{(1/2)} \tag{2.25}$$

The first term in (2.25) corresponds to pure technical efficiency as in RD and FGNZ, the second term is the scale change of FGNZ, the third term is the technical change component of RD and the last term was introduced by the authors and is scale related. This decomposition is based on the FGNZ approach where the authors decomposed the technical change component of FGNZ into a pure change in technology component and a scale change component. This last scale component was called scale bias in Lovell (2001).

Note that irrespective of the approach used to decompose the Malmquist index its value is the same (M_j) for each of the approaches mentioned above. The Malmquist index is, therefore, in all cases calculated through CRS efficiency scores. This is because in order to be a total factor productivity index the Malmquist index should be computed in relation to CRS technologies (e.g. Färe et al. (1997b), Bjurek et al. (1998)). This fact is also mentioned by Grifell-Tatjé and Lovell (1995) who proves that when distance measures are calculated in relation to VRS technologies the resulting Malmquist index mis-measures actual productivity change.

The above decompositions are based on radial (or Farrell type) measures of efficiency, which may be either input oriented or output oriented. In fact the above decompositions are equally valid whatever the orientation used. However, one needs to be aware that efficiency results, and therefore the various components of the Malmquist index, may depend on the orientation of the model, especially when VRS technologies are assumed (under CRS efficiency scores do not depend on the orientation).

The technologies considered above are only two: CRS and VRS. There are, however, other variants in the literature either concerning the technology, and the type of distance measures employed in the computation of Malmquist indexes. For example, Tulkens and Malnero (1996) applied Malmquist indexes in Free Disposal Hull (FDH) technologies, and in Tulkens and Vanden-Eeckaut (1995) the authors go further in using 4 types of technologies (FDH, CRS, VRS, and NIRS) and three types of distance functions: radial input oriented, radial output oriented, and the hyperbolic measure of Färe et al. (1985). Concerning the use of non-radial and non-oriented measures of efficiency to calculate Malmquist indexes the discussion will be postponed to Chapter 8 where we detail on these issues.

2.4 Summary

In this Chapter we have reviewed some concepts on efficiency measurement though data envelopment analysis. DEA implies the specification of a production technology against which efficiency is measured. The technological set can be of various types presenting different returns to scale assumptions or different convexity assumptions. Having specified the technological set, a number of distance measures between observed points and the frontier of the technological set can be specified. Such measures are efficiency measures, which can be of various types ranging from radial to non-radial. In this Chapter we presented some technological specifications and some efficiency measures commonly used in DEA. We also reviewed some other aspects concerning efficiency measurement that will be used to some extent throughout this thesis. Namely, we distinguish between economic and technical measures of efficiency, we point out some models that can be used when there are non-discretionary factors in the data set, we mention some models used to measure super-efficiency and detect influential units, we acknowledge the existence of various forms of restricting the weights in a DEA model, and finally we point out to methods that are usually used to analyse productivity change over time, namely the Malmquist index approach. Each of the basic concepts introduced in this Chapter will be used and/or extended in subsequent chapters of this manuscript.

Chapter 3

Banking Efficiency

In this Chapter we discuss some of the most recent trends in the banking industry, and analyse some important issues involved in the measurement of banking efficiency. Furthermore, we explore some implications of banks being service organisations on the measurement of their efficiency.

3.1 Banking Context

The international financial services industry has been exposed to major changes over the 80's and 90's. Among these, the liberalisation and deregulation of financial markets can be regarded as the leading force of change. By liberalisation and deregulation of financial markets we mean the abolishment of controls over capital movement, interest rates, and credit limits as well as the elimination of most controls and regulations over financial markets (like restrictions to the balance structure, to market access, and to territorial expansion of activities). Developments on information technology (IT) have enhanced and supported most of the changes in the financial markets since the 80's. Indeed, the removal of controls over capital flows allowed free movements of capital, but these would not be effective without an adequate support of IT (Canals, 1995).

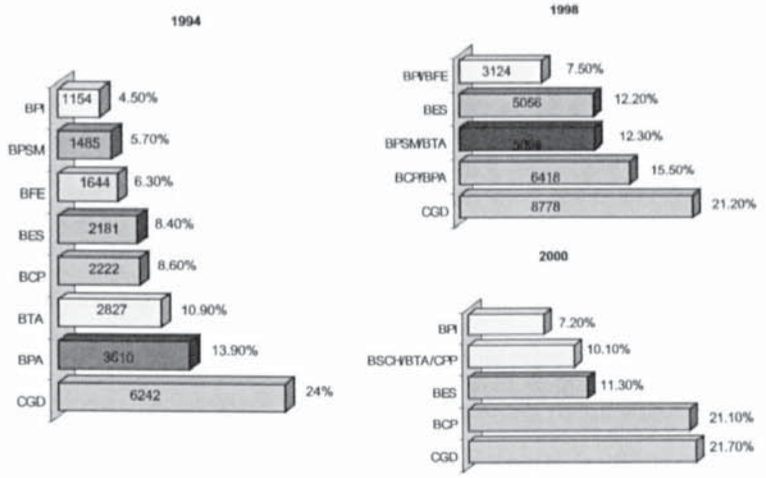
The liberalisation and deregulation of financial markets had two important immediate consequences, (i) an increased internal and external competition, both from financial and from non-financial companies, and (ii) the internationalisation and globalisation of financial activities (e.g. Molineux et al., 1996; Canals, 1995). The increasing competition gave rise to what has been called desintermediation, since banks have lost ground in terms of their intermediation role. Indeed, on the one hand stock markets have attracted customers that previously relied on banks for financing and investing, and on the other hand banks have lost some business for non-financial companies that started offering customers various financing schemes for the acquisition of their products.

The worldwide changes in the financial industry have been particularly felt by European countries due to the European monetary and economic integration. The introduction of the Euro in 01 January 1999 and its physical appearance in 01 January 2002, and all the steps that have been followed towards EMU (Economic and Monetary Union), were important in the process of deregulation and liberalisation of European countries and are still a factor reinforcing the general trends of the industry.

The demand and supply changes on banking services brought about by competition have led to a deterioration of bank profitability and increased their exposure to risk (as the migration of credits from banks to capital markets left banks with higher-risk credits). The responses of banks to this increased competitive pressure passed through mergers and acquisitions, through the diversification of activities (for example entering in new financial business like insurance and stock markets), and through innovation. At the same time cost reductions have been undertaken by banks and efficiency issues have deserved more attention than ever, as under competitive pressures banks cannot afford to waste their resources. It is interesting to note the trend over a rationalisation of resources that banks have been involved with, which has been mainly based on capacity reductions. For the European case the European Central Bank (ECB, 1999b) reports a 26% reduction in the number of credit institutions between 1985 and 1999, and a 7.7% reduction in the number of branches (per 1000 capita) between 1985 and 1997. This decrease in the number of credit institutions reflects mainly the phenomenon of mergers and acquisitions (M&A). In Portugal this phenomenon has also been a reality as can be seen in Figure 3.1. This figure shows that a big percentage of total assets belongs to a small number of banks, and that this number has reduced substantially between 1994 and 2000 (The picture today did not change much from that in 2000).

In spite of the great amount of M&A that took place in Portugal during the 90's, the total number of banks did not decrease (as can be seen in Table 3.1). This is explained on the one hand by the increasing number of foreign banks operating in Portugal, and on the other hand by the fact that mainly acquisitions have occurred and the banks involved did not merge remaining, as such, separate institutions. In Table 3.1 the total number of banks include Portuguese banks and also foreign banks. From 1993 to 2002 the number of

Figure 3.1: Total Assets (billion PTEs) of Larger Banks in Portugal



BPI - Banco Português de Investimento, BPSM - Banco Pinto e Sotto Mayor, BFE - Banco Fomento Exterior, BES - Banco Espirito Santo, BCP - Banco Comercial Português, BTA - Banco Totta e Açores, BPA - Banco Português do Atlântico, CGD - Caixa Geral de Depósitos, BSCH -Banco Santander Central Hispano, CPP - Crédito Predial Português

Table 3.1: Activity Indicators of the Portuguese Banking Sector (Data kindly provided by the Bank of Portugal)

Description	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
Total N.Banks	41	44	45	51	57	59	61	65	61	63
Foreign Banks	7	9	9	12	14	16	17	22	19	19
N. Branches	2872	3115	3489	3725	3979	4159	4467	4668	4781	4823
N. Employees				57531	57347	54975	54738	51128	48637	49490

foreign banks increased from 7 to 19 (about 171%), while the number of Portuguese banks increased at the much slower rate of 29%. In the same period the number of bank branches in Portugal increased by about 68% (see Table 3.1). The increasing trend is consistent over the whole period of analysis, though in recent years the rate of growth has been much lower than in the 90's. In fact from 2000 to 2002 the total number of banks and the number of employees decreased, while the number of bank branches increased at a lower rate. In accordance with what is happening in most EU countries, it is expected that the number of bank branches operating in Portugal eventually starts declining in the near future. The declining trend has been felt more markedly in the number of banking employees that started decreasing after 1996 (see Table 3.1), though in 2002 there was a slight increase in relation to the previous year. The declining number of employees in the banking sector is related, on the one hand, to the search for more efficient ways of working that have been supported by IT developments, and, on the other hand, to the advent of new distribution channels, whose effective use by customers increased immensely in the 90's. This aspect is the subject of the next section.

3.1.1 The Changing Role of Bank Branches

According to ECB (1999a) IT developments can be regarded as the main driving force of change in the banking sector. The influence of IT over banking is both internal and external. The internal changes generated by IT developments consist in the replacement of paper-based and labour intensive methods by automated processes. This substitution makes banks more cost-effective, as the "automation of a significant amount of processing, administration and routine customer service has reduced the branch's overall manpower requirement for traditional bank clerks while simultaneously increasing their output capacity" (Howcroft, 1992, p. 48).

The external changes generated by IT developments resulted in the appearance of a number of new alternative ways by which clients can access banking services. These alternative ways of distributing banking services can be called 'remote banking' (ECB, 1999a) and include phone banking, internet banking, and automatic banking (automatic teller machines - ATMs).

The usage of alternative distribution channels has been growing in the last few years. From the different remote banking channels the least used presently is internet banking. Its use depends on the computer and internet literacy of users, which is not yet very high in Europe. For 1998, ECB (1999a) reports a maximum percentage of internet users of 35% in Finland and a minimum of 1% in Greece. Nordic countries have the highest percentages (33% for Sweden and 22% for Denmark) and Mediterranean countries have the lowest percentages (2% for Portugal, 4% for Italy, 7% for Spain and 6% for France). Concerning phone banking, ECB (1999a) reports that the highest percentage of retail customer's base reached by this distribution channel was 10% (for France and the UK). Concerning ATM machines Table 3.2 (see ECB, 1999a, Table A.I, p. 36) illustrates their rate of growth in recent years. In the five years period considered in this Table the number of ATM's in the EU has grown around 50%. At the same time the number of transactions per capita (last column of Table 3.2) also increased but at a lower rate (46%). Portugal was the country that exhibited one of the largest percentage increase in the number of ATMs (following Greece and Denmark with the highest rates of change), and this change was accomplished by an increase in the number of transactions of around 120%. At the end of 1997 Spain, Portugal, and Austria had the highest number of ATMs per capita (see Table 3.2). Finland is the only country exhibiting a decreasing trend in the number of ATMs, although the number of transactions increased.

Table 3.2: Number of ATM Machines per 1.000.000 Inhabitants

	1993	1994	1995	1996	1997	Change	Transac capita
Belgium	280	313	360	414	492	76%	40%
Denmark	108	142	207	239	253	134%	n.a.
Germany	308	361	437	459	504	64%	n.a.
Greece	82	155	129	185	209	155%	n.a.
Spain	557	600	680	775	863	55%	27%
France	325	355	393	420	462	42%	51%
Ireland	220	241	257	290	286	30%	54%
Italy	262	321	371	422	444	69%	117%
Luxemburg	294	374	456	537	613	109%	5%
Netherlands	292	325	355	373	410	41%	61%
Austria	320	381	420	479	533	67%	38%
Portugal	283	337	372	541	631	123%	120%
Finland	591	557	474	448	445	-25%	8%
Sweden	255	260	266	269	268	5%	24%
UK	328	343	358	376	393	20%	41%
EU	324	363	408	448	488	51%	46%

The increasing number of ATMs has been taking place together with an increasing usage of this distribution channel. This, however, has not been the case with other distribution channels, namely internet and phone banking, whose degree of usage in Portugal is still below expectations. A study carried out by Marktest in 2000 on the Portuguese banking sector[1] reveals that the branch is the distribution channel most preferred by Portuguese customers for almost all the banking services considered except for payments of bills (water, electricity, etc) where ATMs are preferred. Table 3.3 reports some results from the Marktest

Table 3.3: Percentage of Transactions in Different Channels

Description	General Operations	Payment of Bills	Acquisition of Products	Credit	Information on Products held	General Information
Branch	47.7%	41.1%	86.1%	87.8%	43.5%	64%
ATMs	46.2%	48.9%	2.6%	2.8%	40%	14.2%
Phone	2.4%	3%	3.2%	2.9%	7.8%	10.6%
Internet	1.7%	2.8%	2.1%	1.5%	6.1%	7.6%

study, showing the reduced importance of internet in the Portuguese market. For certain type of operations like payment of bills, general operations, and general information on products ATMs play an important role as more than 40% of these operations are done through this distribution channel. The Marktest study further reports that only 4% of customers admitted to use the internet as a means of contact with the bank in the last three months. This same percentage raised to 83.5% for the personal contact at the branch, to 67.5% for contact by means of ATMs, and to 6.7% for phone contacts.

Irrespective of the degree of usage of each alternative distribution channel, there is no doubt that their importance is growing, and that they will be in a very near future the means of choice to perform most banking transactions. This obviously brings in some

[1]Data from this study was kindly provided by the bank, which is the subject of our empirical study.

changes to the traditional way of understanding and undertaking banking activities, which happen both at the bank and at the branch level.

At the bank level new distribution channels bring in a new competitive environment where barriers to entry are likely to be reduced due to the no longer necessary large branch network to reach a "critical mass" of customers. This new competitive environment is also characterised by more informed customers who have easy access to information, which facilitates their mobility between banks. In face of such a dynamic environment banks need to redefine strategies. These are in general shifting from relationship-oriented to product-oriented (Howcroft, 1992, p. 41), which rely less on customer contacts. This obviously is a potential threat to bank branches since an emphasis on products implies its selling through other means than the branch. New distribution channels also bring in a new strategic choice to banks relating with the delivery mix of products. Banks need to choose not only prices or the product mix but also the delivery mix of financial products (see Howcroft, 1992).

At the bank branch level the emergence of new distribution channels represents a threat to their predominance as the means of choice of distributing financial services. This threat might, however, only partly be able to reduce the importance of bank branches, especially due to the increase in personal-advice intensive banking activities that is being undertaken (ECB, 1999b). Indeed, bank branches have been shifting from operating services to consulting (ECB, 1999b), i.e., they are placing less importance on the delivery of transactional services and more importance on exploiting the potential of branch networks as selling outlets for financial services (e.g. Drake and Howcroft, 1995; Howcroft and Beckett, 1993; Howcroft, 1992). In this sense, new distribution channels can be regarded as advantageous to bank branches, since they can place more emphasis on value-added activities (sales related) leaving basic transactions to be performed on other means of distribution. For the bank as a whole this represents a cost advantage, since general operations performed on alternative means of distribution are cheaper than when performed by bank branches' personnel.

The reorientation in the role of bank branches has been supported by layout changes aiming at creating atmospheres more conducive to selling rather than to cheque or savings processing (Howcroft and Beckett, 1993, p. 277). Bank branches are also adopting more pro-active philosophies towards customers. They have been following customers to the places where they go, like shopping centers and supermarkets (Radecki et al., 1996). In these places bank branches have usually extended opening times consistent with the flow of people. At the same time 24 hour banking has become a reality, through ATM machines, call centres, and internet banking that are available 24 hours. Both, layout changes and pro-active attitudes, have been a reality in Portugal and particularly in the bank, which is used in our empirical study.

According to Howcroft (1989) the new distribution channels and IT developments should not be regarded as substitutes of the branch network but as complements. IT can be used to increase the efficiency of bank branches while enabling them to focus on different strategies. As a consequence, despite the growing importance of alternative distribution channels,

> it's not time to give up on the traditional bank branch just yet. After all, it's that familiar sign at street level, and the staff within the branch, that are often a customer's first, and perhaps most comfortable, introduction to the bank, serving as a point of entry to the growing number and diversity of banking products, services and delivery channels. (Howland, 2000, p. 29)

Nevertheless, bank branches survival depends on a refocusing of their activities, and in this process the measurement and monitoring of their efficiency assumes particular relevance.

3.2 Traditional Methodologies in Banking Efficiency Assessments

The methodologies that have been used to assess banking efficiency can be classified into three main strands. The oldest strand relates to studies using econometric cost functions to analyse scale and scope economies in banking. These studies assume that production units are efficient, which is consistent with traditional economic theory. The second strand is also econometric or stochastic but uses frontier methods and aims at assessing not only scale and scope economies but also (in)efficiency. The third and most recent strand also uses frontier methods but these are deterministic or non-parametric. Both parametric and non-parametric frontier methods were inspired by the work of Farrell (1957), although the roots of parametric techniques go back to 1928 with the work of Cobb and Douglas who developed a production function that is known by their names (Greene, 1993, p. 68). Non-parametric frontier methods have been mostly used after its operationalisation by Charnes et al. (1978) that gave rise to Data Envelopment Analysis (DEA).

Parametric techniques can be classified into three categories [according to Berger and Humphrey (1997)]: (i) the Stochastic Frontier Approach (SFA), (ii) the Thick Frontier Approach (TFA), and (iii) the Distribution Free Approach (DFA). Each of the stochastic frontier methods has advantages and disadvantages. In the article where he develops the DFA, Berger (1993) criticises the two other methods, referring to their strong and unverified assumptions on the statistical properties of efficiency and random error terms. In SFA the use of the half normal distribution for the inefficiency term, is, according to Berger (1993) inflexible relative to other distributions and "it embodies the arbitrary restriction that most firms are clustered near full efficiency" (Berger, 1993, p. 262). For the case of TFA the main criticism of Berger (1993) concerns the arbitrariness of where inefficiencies stop and random error begins. According to Berger (1993) the DFA has the advantage of not imposing such arbitrary assumptions on the data. Its main disadvantage is, however, the fact that only one average efficiency value is provided by the method. That is, the use of panel data not allow the analysis of efficiency changes over time since the final result is a single average over time efficiency measure for each unit.

In stochastic frontier methods two choices need to be made: one concerning the method for drawing the efficient boundary (which can be one of the three above), and another concerning the type of functional form to use. According to Berger and Mester (2000) the translog is the most popular form in the literature. However, this functional form has some problems and a more flexible form, the Fourier-flexible functional form, has been recently preferred by some authors.

In summary there are some choices to be made when efficiency is to be measured through frontier methods. First one needs to decide whether to use parametric or non-parametric techniques, and then within each technique a number of other choices are to be made. Doubts naturally arise concerning whether parametric methods are better than non-parametric ones, or whether a given specification is better than another inside the same general method. According to Berger and Humphrey (1997) there seems to exist no agreement concerning parametric vs non-parametric methods, as all methods have advantages and disadvantages. Among these, Berger and Humphrey (1997) refer to the fact that parametric methods have the disadvantage of imposing a functional form to the *best-practice* frontier and demanding some assumptions regarding the behaviour of the parameters of the model. If the functional form specified is not the correct one, then efficiency will be confounded with specification errors[2]. On the other hand non-parametric methods do not impose any functional form on the frontier[3], but have the disadvantage of not allowing for random errors due to chance, data problems, or measurement errors. If this random error exists then it will be entangled with actual inefficiency. An advantage of non-parametric methods is also the fact that they result in individual efficiency measure for each observation (Førsund et al., 1980).

As the true level of (in)efficiency is unknown it is impossible to state without doubts whether parametric techniques outperform non-parametric techniques or vice-versa. Empirical evidence suggests, however, that the two sets of methods give similar results in terms of efficiency scores, although they tend to disagree in terms of variation and ranking. For example, in their literature review, Berger and Humphrey (1997) compared average efficiency for banks under the two broader classes of methods and reached similar values of efficiency. Berger and Humphrey (1997) report a similar average efficiency for the two methods in US banks with parametric methods providing higher estimates of efficiency, and lower dispersion on estimated efficiency (standard deviation of 0.17 for non-parametric techniques against a value of 0.06 for parametric techniques). Ferrier and Lovell (1990) also compared the two methods (DEA and stochastic frontier methods) in an application to 575 banks in the US, and concluded that "the two approaches are in substantial agreement on several important issues" (see p. 24).

The reference to differences in ranking is not supported by strong empirical evidence as only a few banking studies compare ranks across the two techniques. The study by Ferrier and Lovell (1990) reported a positive correlation between rankings, but this was not statistically significant. Giokas (1991) compared DEA and the loglinear estimation function using 17 bank branches in Greece. The models provided similar results concerning the significance of certain variables, but there were some differences concerning the classification of units. Under DEA some units had an efficiency score of 100% whereas under the loglinear model these same units had an efficiency score significantly below unity. Rankings may be different when comparing parametric and non-parametric methodologies, but there are also significant ranking differences within the same class of method. For example, Bauer et al. (1993) compared, in the banking context, SFA with TFA and concluded that whereas the

[2] As referred to previously these methods, in particular the SFA, may have other problems associated with the distributional assumptions imposed on the inefficiency and error terms (see Berger, 1993).

[3] There is, however, some structure embodied in the assumptions that underly the construction of the efficient boundary.

two approaches yielded similar results in terms of inefficiencies "they rank individual banks quite differently" (see p. 410).

Outside the banking literature comparisons between parametric and non parametric methodologies have also been undertaken. For example Gong and Sickles (1992) compared DEA with some stochastic frontier methods applied to a simulated data set where the true production frontier was known. The authors concluded that stochastic frontier methods outperform DEA when the functional form chosen was close to the underlying technology, being the reverse true when the functional form was not close to the underlying technology. In real world applications it is difficult, if not impossible, to know whether the functional form chosen resembles the underlying technology, which gives an advantage to DEA over stochastic methods. In a number of studies by Banker and colleagues (Banker et al., 1986, 1988, 1993, 1996b) DEA was compared with parametric methods in a number of respects. Conclusions point out the superiority of DEA, whose accuracy increases with the size of the sample of observations. Hjalmarsson et al. (1996) compared two deterministic methods with stochastic frontier methods in an application to the Colombian cement industry. In general the authors conclude that "efficiency scores and the time pattern of efficiency vary as much within each class of models as across models" (see p. 321).

In summary, the above studies suggest that more important than choosing the technique to be employed in a particular data set is the choice of the model specifications within that technique. In addition, the advantages/disadvantages of parametric and non-parametric techniques are complementary which suggests that the use of both methodologies can bring some light to the understanding of the production process being modelled.

3.3 Different Efficiency Approaches - The choice of Inputs and Outputs

As banks and bank branches are complex units of analysis it is almost impossible to analyse their efficiency without focusing on specific aspects of the production process. An attempt to include all inputs and outputs that banks or branches consume and produce might prove quite difficult. On the one hand the choice of input and output factors might be a problematic issue, and on the other hand the expected high number of such factors might turn the analysis, parametric or not, of limited significance. For this reason authors have focused on some perspectives of the activity of banks and bank branches to measure their efficiency. The main approaches one can distinguish are the *production* and *intermediation* approaches to banking activities (see Colwell and Davis (1992) for a detailed discussion on both approaches).

The **production approach** takes the perspective of operations that go on in a bank or bank branch. These are seen as production units that use a set of resources, such as staff, computer terminals, space, etc. to produce a set of services such as various types of accounts, loans, insurance, etc. In this sense banks are seen as service providers, where the emphasis is on operational activities rather than on market oriented financial activities. Under this approach inputs relate to personnel and capital, and outputs are usually the number of transactions performed at the bank or the number of accounts of various types held by the bank. Interest costs and revenues are excluded from this approach. Inputs and outputs should, under the production perspective, be measured in physical units (Berger and Humphrey, 1997). Most of the existing studies in the literature that focus on *production efficiency* use the bank branch as the unit of analysis [e.g. Sherman and Gold (1985), Athanassopoulos (1997), Vassiloglou and Giokas (1990), Giokas (1991), Drake and Howcroft (1994), Schaffnit et al. (1997), and Camanho and Dyson (1999)]. Some studies that focus on banks and use this approach can be seen in Ferrier and Lovell (1990), Berger et al. (1987), and Berg et al. (1991).

The **intermediation** approach looks at banks or bank branches as intermediaries - collectors of funds that are intermediated into loans and other assets. Under this approach, and in simple terms, the bank borrows funds at some price and sells them at a different and higher price. In this sense the bank earns some revenues which are the outputs of the efficiency assessment. It also incurs some costs, which are considered the inputs of the assessment. The inputs that are most commonly used in the literature under this approach are the interest and non-interest costs (service rates, commissions, salaries and other expenses). The outputs are usually interest income and non-interest income, total volume of loans, deposit accounts, saving deposit accounts, etc. Most of the existing studies in the literature that focus on this approach use banks as the unit of analysis. Some examples can be found in Berger and Mester (2000), Berger et al. (1987), Berger and Humphrey (1991), Aly et al. (1990), Barr et al. (1993), etc. Bank branch's studies using the intermediation approach can be seen in Berger et al. (1997) and Athanassopoulos (1997).

Both the intermediation and the production approaches obviously fail to reflect each one on its own the global set of activities going on at a bank or its branches. They need to be used simultaneously if one wants to get a complete picture of the bank or branch efficiency (see for example Athanassopoulos (1997) or Berger et al. (1997) who used both approaches). Berger and Humphrey (1997) and Thanassoulis (1999a) point out that financial institution have a dual role of providing transaction services, and of acting as financial intermediaries. In this sense both the intermediation and production approaches should be used together since they respond to different issues. The production approach is appropriate to measure cost efficiency as it uses mainly variables relating to operational costs of banks, and the intermediation approach, which is concerned with overall costs, is more appropriate to answering questions related to the economical viability of banks (Ferrier and Lovell, 1990, pp. 230-231). According to Berger et al. (1997) the intermediation approach is generally preferred for analysis at the bank level because it is more inclusive and it captures the essence of a bank as a financial intermediary. At the same time substitutions between operating costs and interest costs (increasing one while decreasing the other) are only captured by the intermediation approach (Berger et al., 1997). According to the same author, for an analysis at the branch level the production approach might be more appropriate. This is because the asset and liability decisions are taken by banks as a whole, while branches are primarily producers of depositor services.

Within each of the two broad approaches to analyse banks or bank branch's activities there is reasonable agreement about the factors that should be considered inputs or outputs of the efficiency assessment. The main controversial factor is bank's deposits, which can be both considered as an input (as banks receive money from deposits, which is then used to create other banking assets) and as an output (as they are one amongst a set of services banks provide to their customers) (see e.g. Berg et al. (1991), Colwell and Davis (1992), Adams et al. (1999) or Berger and Humphrey (1997) for a discussion on this subject). There are three approaches to help assigning banking factors to the input or output categories. These approaches are:

> (1) the asset approach (only assets are outputs), (2) the user-cost approach (accounts that provide net revenue above the opportunity costs are outputs), and (3) the value-added approach (accounts associated with large expenditures of real resources are outputs). (Adams et al., 1999)

According to each of these approaches deposits may be considered inputs or outputs. In the asset approach they are inputs, and in the value-added approach they are usually considered outputs. In the user-cost approach they may be considered either inputs or outputs (Adams et al., 1999).

In the study by Adams et al. (1999) the authors tested various model specifications assuming in all the cases that loans are outputs, but changing the specification of deposits and savings either as inputs or outputs. The results suggested that the specification of deposits as inputs statistically dominated the other model specifications. Tortosa-Ausina (2002) also analysed different output specifications considering deposits in one case both on the input and output side, and in the other case only on the input side of the DEA assessment. Results point out for great differences in the shape of the efficiency distribution depending on the output specification. This suggests that it is not indifferent to consider deposits on the input or output side of the efficiency assessment. One way to avoid such an ambiguity is to use a dual approach capturing both the input and output characteristics of deposits (Berger and Humphrey, 1997). This was done in Berger and Humphrey (1991) who considered quantity of deposits on the output side of the efficiency assessment, and interest paid on deposits on the input side of the assessment.

Another discussion in the literature concerns the units of measurement of the factors used in the analysis. The units of measurement can either be represented in quantity or value of accounts. Typically the production approach uses number of accounts as outputs, and the intermediation approach uses value of accounts as outputs. Arguments in favour of using number of accounts include the fact that providing accounts or credit to clients results in expenses that are not directly related with the value deposited or loaned. Arguments in favour of using the value of accounts include the fact that the market evaluates a bank in terms of value and not in terms of number of accounts (Molineux et al., 1996, Chapter 5).

In the study of Berg et al. (1991) the importance of selecting appropriate units of measurement in an efficiency analysis was discussed. The authors used DEA to evaluate technical, and scale efficiency considering two different output measures, one expressed in monetary terms and the other considering the number of accounts and their average size. Results suggested that there were not many differences between using one or the other type of output measure as far as the characteristics of the efficient frontier were concerned. As for the ranking of individual banks this was affected by the output measure used, meaning that "before undertaking efficiency studies, one should therefore take great care to clarify what kind of efficiency one wants to measure, and define output measures accordingly" (Berg et al., 1991, p. 140).

3.4 Banks as Service Organisations: Implications

"Services are 'different' and as such they challenge us to reformulate or develop new models to analyze them" (Chase and Heskett, 1995). For this reason it is important to analyse the implications of banks being a service organisation in measuring their efficiency.

As service organisations banks present a number of characteristics specific to services. The differentiating characteristics of services are mainly four: intangibility, perishability, simultaneity of production and consumption, and heterogeneity (e.g. Klassen et al., 1998; Lewis, 1989). The intangible nature of services makes their quantification difficult. The simultaneous consumption and production and perishable nature of services makes it difficult to determine their production capacity in the absence of immediate demand. In addition, the simultaneity of services makes it difficult to disaggregate marketing and production functions, as these are often undertaken by the same person. As a result, the participation of the customer in the service influences the production process and may have an effect on the quality perceptions of customers. Services are also, in their intrinsic nature, heterogeneous as each customer requires in one way or another a specific way of being served [for details on the impact of service characteristics see Lewis (1989)].

A direct result of banks being service organisations is that efficiency is usually harder to measure (and define), and is linked with a set of other important performance measures such as service quality and effectiveness (Lovelock, 1996, Chap. 11). Let us consider a simple manufacturing organisation that produces chocolate cakes and define its efficiency as the extent to which the maximum output is produced given the set of inputs used. If we use a single measure of output which is the quantity of cakes produced, and if inputs are the labour hours used to produce the cakes, the ingredients used, and the technology employed, then a ratio of the outputs produced per inputs used is a pure internal measure of efficiency. Now suppose that instead of cakes produced we use as a measure of output the number of cakes sold. In such a measure of efficiency we would be integrating external factors related to the ability of the product to sell in the market place[4]. This ability is strongly dependent, amongst other factors, on the quality of the product. More tasty cakes will probably sell better and the output of firms producing these cakes (when output is measured through sales) will be higher. Also the effectiveness of the firm is included in this

[4]Achabal et al. (1984) mentioned this aspect as far as productivity measurement is concerned. According to the author it is essential to distinguish between the actual output and the sales of that product, which might be especially complicated in service industries (see also Klassen et al., 1998).

measure of efficiency as, in a certain sense, we are also measuring the extent to which the firm is meeting its objectives (to sell the cakes produced).

Traditionally efficiency and productivity are concepts geared to manufacturing organisations because they assume that production and consumption are separate processes and that the customer not participate in the production process (Chase and Heskett, 1995; Grönroos, 2000). However, in service organisations, like banks, outputs do not correspond to things that were produced and not yet sold. Outputs are in most cases things that customers already bought and are more or less happy with. Lovelock (1996) refers in relation to this aspect that the outputs of service organisations are not outputs in the manufacturing sense but outcomes. For this reason, the measurement of efficiency in banking or other service organisations implicitly incorporates a service quality dimension, especially when we understand this measurement in the long run. This is the perspective of Lovelock (1996, p. 468) who mentions that "the need to emphasise effectiveness and outcomes suggest that issues of productivity cannot be divorced from those of quality and value".

Efficiency is, therefore, usually considered an internal concept measured according to the "constant-quality assumption" (*changes in production inputs do not have an impact on the quality produced* (Grönroos, 2000, p. 206)). This means that in manufacturing, as long as outputs are the same, customers are willing to buy this output even if it was produced using a different input mix. In services this might not be so. The purchase intentions of the customer, given an altered input mix, will depend on the perceived "process-related and outcome related quality of the new resources or inputs used" (Grönroos, 2000, p. 209).

Therefore, in services the "constant-quality assumption" not apply and as such measurement of efficiency cannot be disentangled from service quality. Grönroos (2000) distinguishes between two concepts of efficiency: internal efficiency and external efficiency. The concept of internal efficiency is related to the efficiency by which outputs can be produced through a given amount of inputs, and external efficiency is related to the efficiency by which perceived service quality is produced through a given amount of inputs. It is of no use to improve internal efficiency if external efficiency is not taken into account. That is, in service organisations the introduction of more "cost effective production resources and processes does not necessarily lead to better economic results" (Grönroos, 2000, p. 206). This is because internal efficiency is related to cost efficiency (cost savings can be achieved by improving internal efficiency) but external efficiency is related to revenue efficiency (decreasing service quality results in lost revenues). This view is shared by other authors like

Howcroft (1991) and Thanassoulis et al. (1995, p. 588) who mentioned that "a unit running at high levels of output but of low quality may be efficient in providing output volume for its resource level. However, by providing low quality outputs it could lose viability through loss of customers or face outside regulatory intervention".

While in manufacturing organisations it may be easy to disentangle and measure separately internal and external efficiency, in service organisations this is not so. For this reason, both internal and external efficiency need to be accounted for in measuring the efficiency of service organisations. If the former is related with cost and the latter is related with revenue then an obvious way of linking both is through profit. The measurement of profit efficiency appears therefore as a way to incorporate both cost efficient practices and a quality dimension that is likely to generate revenues. This view is shared by the service marketing literature, as seen above, and by the banking efficiency literature as testifies the work of Berger et al. (1993). According to this author, profit efficiency measures can be used as a means to account for differences in quality between banks. As high quality may be produced at the expense of higher costs, a higher quality bank or bank branch could be deemed cost inefficient if the revenues generated by increased quality were not accounted for in the analysis (see also De Young and Nolle, 1996; Berger and Mester, 2000). Two different strands of the literature arise, therefore, at a similar conclusion: that a profit efficiency analysis takes into account both internal (cost related) and external (revenue related) aspects of banks' activities being the best way to thoroughly measure efficiency in the case of service organisations.

3.5 Summary

The aim of this chapter was to put into context the measurement of bank branches' efficiency, the main subject of this manuscript. For that purpose we first point out the main changes that the banking industry has been facing over the past few years, and acknowledge the changing role of bank branches within this context. We also analyse in broad terms the main issues that arise when banking efficiency assessments are undertaken, namely the main type of methodologies at the disposal of the researcher, and issues relating with the way the production process is approached. Finally we note that banks and its branches are service organisations and mention the implications of this fact on the measurement of their efficiency.

Chapter 4

Literature Review on Bank Branches Efficiency Measurement

The previous Chapter has introduced some general issues in measuring efficiency in banking. In this Chapter we review the literature on bank branch efficiency studies focusing particularly on the issues that are relevant for the development of our own models to assess the efficiency of a sample of bank branches. These issues relate to the extent to which the new role of bank branches has been acknowledged and dealt with in extant studies, and the way these studies have linked efficiency with service quality and profitability.

In the banking literature there are studies that analyse banks as a whole, and studies that analyse bank branches of a single bank. Our focus on the latter strand of the banking literature is justified by our empirical application to bank branches.

4.1 Traditional Measurement in Bank Branches

The traditional technique used by banks to compare the performance of its branches is performance ratio analysis where a set of ratios are computed to capture different dimensions of the banking process (Schaffnit et al., 1997). Such ratios can be, for example, return on assets, return on investment (Sherman and Gold, 1985; Oral and Yolalan, 1990), profit per average investment (Sherman and Gold, 1985), transactions per teller, and cost per transaction (Sherman and Ladino, 1995).

Financial ratios for comparing bank branches are more common than non-financial ratios, which means that dimensions such as productivity and quality are neglected in these comparisons (Schaffnit et al., 1997). In addition, ratios fail to account for the interactions and trade-offs between different factors in the process, therefore providing an incomplete picture of that process (Schaffnit et al., 1997). Ratio analysis often results in a set of contradictory measures that show good performance of branches on some dimensions and bad performance on other dimensions. Clearly decisions based on such contradictory information are difficult to make.

Besides these disadvantages, ratios are largely used due to their simplicity of calculation and interpretation. However, given their limitations ratio analyses should be complemented with other operating measures such that a fair comparison between bank branches is undertaken. Efficiency analysis of bank branches through both parametric and non-parametric techniques has been one of the major tools used in the most recent literature to reach a fair comparison of bank branches. These bank branch efficiency studies are reviewed in the next section.

4.2 Efficiency Studies on Bank Branches

In Table 4.1 we show published studies analysing bank branches' efficiency. The earliest study in this Table dates back to 1982 and is due to Murphy and Orgler (1982). The aim of the study of Murphy and Orgler (1982) is not, however, to compare bank branches regarding their efficiency, but rather to analyse the determinants of bank branches' costs and scale economies. In fact most of the parametric studies included in Table 4.1, such as Pavlopoulos and Kouzelis (1989), Doukas and Switzer (1991), and Zardkoohi and Kolari (1994) aim at analysing the degree and type of scale and/or scope economies applying at bank branches, rather than their efficiency[1].

We can, therefore, consider the study of Sherman and Gold (1985) as being the first published study to analyse the efficiency of bank branches. In an earlier study Sherman (1984) already referred to the assessment of bank branches through DEA, but did not apply the methodology.

Since 1985 a relatively small set of articles analysing the efficiency of bank branches has been published. From published studies there is a clear dominance of non-parametric methodologies over parametric ones to analysisng bank branches' efficiency (see Table 4.1).

[1]Note, however, that these issues are closely related with the concept of efficiency and therefore we included these earlier studies in Table 4.1.

Table 4.1: Bank branch studies

ID	Study	N. Units	Country	Models	Orientation
		Non-Parametric Approaches			
1	Sherman and Gold (1985)	14	US	CCR	Input
2	Parkan (1987)	35	Canada	CCR	Input
3	Oral and Yolalan (1990)	20	Turkey	CCR	Input
4	Vassiloglou and Giokas (1990)	20	Greece	CCR	Input
5	Giokas (1991)	17	Greece	CCR & BCC	Input
6	Oral et al. (1992)	44	Turkey	CCR	Input
7	Al-Faraj et al. (1993)	15	Saudi Arabia	CCR	Input
8	Tulkens (1993)	> 700	Belgium	FDH & CCR & BCC	Input
9	Drake and Howcroft (1994)	190	UK	CCR& BCC	Input
10	Sherman and Ladino (1995)	33	US	CCR	Input
11	Haag and Jaska (1995)	14	US	Additive	Non-oriented
12	Tulkens and Malnero (1996)	663	Belgium	FDH	Input
13	Nash and Karwat-Sterna (1996)	75	?	Additive	Output
14	Soteriou and Stavrinides (1997)	26	Cyprus	CCR	Inp & Out
15	Athanassopoulos (1997)	68	Greece	CCR & non-radial	Inp & non-oriented
16	Schaffnit et al. (1997)	291	Canada	CCR & BCC	Input
17	Lovell and Pastor (1997)	545	Spain	BCC	Output
18	Athanassopoulos (1998)	580	UK	CCR & BCC	Inp & Out
19	Camanho and Dyson (1999)	168	Portugal	CCR & BCC	Inp & Out
20	Zenios et al. (1999)	144	Cyprus	CCR	Input
21	Soteriou and Zenios (1999)	144	Cyprus	BCC	Inp & Out
22	Golany and Storbeck (1999)	182	US	BCC	Output
23	Kantor and Maital (1999)	250	?	CCR?	Output
24	Avkiran (1999a)	65	?	CCR& BCC	Output
25	Soteriou and Stavrinides (2000)	26	Cyprus	CCR	Inp & Out
26	Athanassopoulos (2000)	60	Greece	Non-radial	Out & non-oriented
27	Athanassopoulos et al. (2000)	126, 185, 196	Cyprus& Greece&UK	CCR & BCC	Inp & Out
28	Athanassopoulos and Giokas (2000)	47	Greece	CCR	Inp & Out
29	Cook et al. (2000)	20	Canada	Specific	Input
30	Cook and Hababou (2001)	20	Canada	Additive	Inp & Out
31	Dekker and Post (2001)	314	Holland	quasi-concave DEA	—
32	Hartman et al. (2001)	50	Sweeden	BCC	Input
		Parametric Approaches			
33	Murphy and Orgler (1982)	127	?	Cobb Douglas	Cost
34	Pavlopoulos and Kouzelis (1989)	362	Greece	Translog flexible cost function	Cost
35	Doukas and Switzer (1991)	563	Canada	Translog cost function	Cost
36	Zardkoohi and Kolari (1994)	615	Finland	Translog cost function	Cost
37	Boufounou (1995)	62	Cyprus	Regression	—
38	Kamakura et al. (1996)	188	Latin America	Translog cost function	Cost
39	Berger et al. (1997)	560	US	DFA	Cost
40	Avkiran (1997)	93-115	Australia	Regression	—

Our analysis of the studies in Table 4.1 will first focus on general aspects, like the type of data used, and the type of models used. Secondly, we will review those studies that have incorporated quality issues into the efficiency analysis, and studies that have incorporated the profitability dimension into the analysis. We will not analyse and compare specific results from the various studies because, as it is clear from Table 4.1, these concern quite distinct countries, sample sizes, and models. In addition, the set of inputs and outputs used in each case is so different that any attempt to compare efficiency results is of reduced interest.

4.2.1 The Data

Cross Sectional Vs Panel Data

Most of the studies in Table 4.1 used cross-sectional data rather than panel data, meaning that the analysis undertaken focused on data from a single period. The exceptions in Table 4.1 are Golany and Storbeck (1999) who used six quarters of data, Tulkens and Malnero (1996) who used eleven months of data, Hartman et al. (2001) who used two years of data, and Berger et al. (1997) who used three years of data. Although multiple year data sets have been extensively used in banks' efficiency assessments, bank branches' assessments have mainly focused on single period analysis. This is a clear limitation of these studies as efficiency is an evolving concept and decision makers naturally have the interest to analyse efficiency evolution over time. As Golany and Storbeck (1999, p. 19) mentioned, "observing the results for a single period rarely leads to decisions in any real world application".

The studies of Golany and Storbeck (1999), Berger et al. (1997), and Hartman et al. (2001) using multiple period data performed the analysis independently for each period. This means that production frontiers were not compared between periods and no technical progress/regress was assessed in these cases. That is, the assessments assumed that technology did not change from period to period, and as such changes in efficiency were not attributed to any other cause than managerial practices. In the study of Tulkens and Malnero (1996) the authors accounted for technological change by calculating Malmquist indexes to evaluate frontier shifts and efficiency change over time. These indexes were calculated for FDH type frontiers following the procedure detailed in Tulkens and Vanden-Eeckaut (1995).

Inputs and Outputs

Most of the studies in Table 4.1 adopt the production approach to analyse bank branches efficiency (Sherman and Gold, 1985; Parkan, 1987; Vassiloglou and Giokas, 1990; Giokas, 1991; Tulkens, 1993; Drake and Howcroft, 1994; Sherman and Ladino, 1995; Athanassopoulos, 1997; Schaffnit et al., 1997; Camanho and Dyson, 1999; Zenios et al., 1999; Soteriou and Zenios, 1999; Golany and Storbeck, 1999; Avkiran, 1999a; Athanassopoulos et al., 2000; Athanassopoulos and Giokas, 2000; Hartman et al., 2001). The studies that use an intermediation approach are those of Oral and Yolalan (1990), Zardkoohi and Kolari (1994), Berger et al. (1997), Athanassopoulos (1997), and Athanassopoulos (2000). Apart from intermediation and production approaches, there are studies that introduce new measures of efficiency like service efficiency (Oral and Yolalan, 1990; Oral et al., 1992), market efficiency (Athanassopoulos, 1998), and quality efficiency (Soteriou and Stavrinides, 1997; Soteriou and Zenios, 1999; Soteriou and Stavrinides, 2000; Athanassopoulos, 2000). In addition, some studies in Table 4.1 analyse bank branches with respect to other issues like sensitivity analysis of DEA results (Haag and Jaska, 1995), the methodology of treatment of panel data (Tulkens and Malnero, 1996), the analysis of branch cross-selling efficiency (Nash and Karwat-Sterna, 1996), the ability of bank branches to meet management established targets (Lovell and Pastor, 1997), the integration of DEA and ABC (activity-based costing) analysis (Kantor and Maital, 1999), the analysis of the determinants of banking costs and the investigation of scale economies (Murphy and Orgler, 1982; Pavlopoulos and Kouzelis, 1989; Doukas and Switzer, 1991), and the analysis of branch performance determinants (Boufounou, 1995; Avkiran, 1997).

Irrespective of the approach used to analyse bank branches' efficiency a set of inputs and outputs is chosen for this purpose. It is of interest to analyse the main types of inputs and outputs that have been used to assess bank branches' efficiency under the two broad production and intermediation approaches so that our own choice of input and output factors to be detailed in Chapter 5 is contextualised.

In Table 4.2 we show the main inputs and outputs that have been used in production and intermediation approaches. In this table we also identify the studies which have used each input/output through their identification number (ID) shown in Table 4.1.

Table 4.2: Inputs and Outputs used in Bank Branch Studies

Inputs	Outputs
Production Approach	
Staff related	**Transactions related**
Total N. Staff (1, 2, 3, 6, 7, 8, 11, 15, 19, 24, 32)	Total n. transactions (2, 19)
N. staff grouped by type (10, 16, 20, 21, 29, 30)	N. transactions grouped by complexity or time required (1, 4,5, 10, 11 28)
Personnel hours (4, 5, 12, 28)	N. transactions grouped by other criteria (8, 9, 12, 15, 16, 28, 29, 30)
Staff Grades (9)	N. transactions in external ATMs (8, 19)
Percentage employees with college degree (7)	Time spent on various types of transactions (3, 6, 21)
Average years of experience (7)	N. account openings (2, 39)
N. teller hours (22)	N. loan applications (2, 39)
N. non-teller hours (22)	N. accounts closed (39)
Staff cost (7, 27)	**N. Accounts**
Technology related	N. accounts (19)
N. Computers (2, 3, 4, 6, 15, 20, 21, 27, 28, 32)	N. term accounts (16)
N. ATMs (8, 9, 15, 19)	N. current accounts (7, 27)
N. windows operated (8, 24)	N. deposit accounts (15, 39)
N. interview rooms (9)	N. savings accounts (27)
N. teller machines (15)	N. business accounts (27)
Branch Space (m2) (4, 5, 9, 10, 19, 20, 21, 22, 27, 28, 32)	N. loan accounts (16, 27)
Rent (1, 2, 11, 39)	N. debits (39)
Quality variables	N. credits (39)
Quality of customer service space ranking (2)	N. new deposits accounts (24)
Staff conduct (24)	N. new lending accounts (24)
Marketing and Environmental variables	**Value/balance of Accounts**
Location index (7)	Value of savings (19)
Marketing activity ranking (2)	Value of loans (19, 22, 32)
N. commercial accounts (3, 6)	Value of deposits (22, 32)
N. saving accounts (3, 6, 20, 21)	Value of mortgages (7, 32)
N. credit applications (3, 6, 20, 21)	Value of current accounts (7)
N. current accounts (6, 20, 21)	Value of saving accounts (7)
Commissions (21)	Value of other accounts (7)
Foreign currency and commercial accounts (20, 21)	**Quality Variables**
Mailing expense per customer (22)	Customer service survey rating (2)
Unemployment statistic (22)	Number of corrections (2)
Average annual family income (24)	Depth of relations (22)
Presence of competitors (24)	Satisfaction (22)
Other Variables	**Other Variables**
Supply/operational Costs (1, 2, 4, 5, 7, 9, 10,11, 19, 28, 39)	Average profit per month (7)
Wages price (39)	Loan index (7)
Index for expenditure on decoration (7)	N. customers (32)
	Fee income (24)
Intermediation Approach	
Personnel expenses (3, 6)	Interest earned on loans (3, 6)
Administrative expenses (3, 6)	Non-interest income (3, 6, 15)
Depreciation (3, 6)	Commissions (26)
Interest paid on deposits (3, 6)	Value of loans (15, 36)
Total interest costs (15)	Value of time deposit accounts (15, 26)
Non-interest costs (6, 15, 26)	Value of saving deposit accounts (15, 26)
Total costs (operating + interest expenses) (36, 39)	Value of current deposit accounts (15)
Capital price (Cost per m^2 of office space) (36, 39)	Value of demand deposits (26)
Wage price (36, 39)	Value of retail deposits (36)
Deposits price (interest rate paid on deposits)(36)	Value of consumer transactions accounts (39)
Branch Size (26)	Value of consumer non-transactions accounts (39)
Convenience (26)	Value of business transactions accounts (39)
Mean queue length (26)	Value of business non-transactions accounts (39)

When a production approach is used to measure bank branches' efficiency the inputs chosen relate mainly to the work force, space of the branch, and supply or operational costs. The work force is either considered in number of staff, hours spent by staff, staff costs, or staff grades (see Table 4.2). In some studies personnel is also broken down by their function (for example, tellers, platform staff, and managers are used in Sherman and Ladino (1995); and tellers, ledgers, accounting officers, typing staff, supervision personnel, and credit staff are used in Schaffnit et al. (1997)). Technological variables (or capital related variables) relate mainly to the space of the bank branch (usually measured in square meters), number of computers and number of ATMs. In some studies space costs are proxied by the variable rent. This variable may also be used to reflect the location of the bank branch, in the sense that central bank branches in general pay a higher rent than rural bank branches.

Less common inputs that have been used in the literature include quality variables used by Parkan (1987), marketing and environmental variables used by Golany and Storbeck (1999), and variables reflecting the *micro-environment* of the branch used by Oral and Yolalan (1990), Soteriou and Zenios (1999), and Zenios et al. (1999). Note that the variables used to reflect the *micro-environment* of the branch are the number of various types of accounts, which are usually considered outputs of production efficiency assessments. The authors argue that these variables "reflect the steady-state market conditions the particular branch has reached due to its previous efforts" (Zenios et al., 1999, p. 42).

Concerning outputs, a production efficiency assessment typically uses the number of transactions as the main output of the model. Transactions performed in a bank branch are numerous and have different complexity levels requiring, therefore, different amounts of staff time. For that reason, transactions are usually grouped according to their complexity, the time required to perform the transaction, or other criteria. In most cases transactions include not only basic activities like deposits or withdrawals, but also sales of deposit and credit accounts. In this sense new and closed accounts are included in one of the transactions type (e.g. Sherman and Gold (1985) and Sherman and Ladino (1995)) as well as loans applications and payments (Sherman and Gold, 1985; Sherman and Ladino, 1995; Athanassopoulos, 1997). The number of transactions is preferred to the number of accounts to measure the output of bank branches, given the full and unlimited access of customers to all branches of a given bank. This means that it is not the number of accounts of a given bank branch that determines the operational work load of that branch, but the number of transactions that are performed there each day.

In some studies transactions are used alongside with other outputs. For example, Schaffnit et al. (1997) used, apart from transactions, the number of term accounts, the number of personal loan accounts, and the number of commercial loan accounts. According to the authors, number of accounts was included as a *proxy* for the work spent by branch personnel on maintenance. Camanho and Dyson (1999) used not only the number of general service transactions on the output side, but also the number of accounts, the number of transactions in external ATMs, the value of savings, and the value of loans. Berger et al. (1997) also used the number of deposit accounts, the number of debits, and the number of credits, together with transaction related variables (namely the number of accounts opened, accounts closed, and loans originated).

There are also some studies in the literature that did not use transactions on the output side of production efficiency assessments. Oral and Yolalan (1990), for example, used outputs relating to the time spent in four types of services (general service transactions, credit transactions, deposit transactions, and foreign exchange transactions). This was also the first study to include time variables in the analysis, which according to Schaffnit et al. (1997) provides useful information on the production process (see also Soteriou and Zenios, 1999; Zenios et al., 1999). Athanassopoulos et al. (2000) uses only number of accounts on the output side, namely the number of current accounts, of savings accounts, of company accounts, and of credit accounts. In the study of Golany and Storbeck (1999) the outputs considered were the volume of loans and the volume of deposits, and also two outputs intending to capture a quality dimension of the service. The authors regarded the outputs on service quality as representing long term objectives, contrary to the first two outputs that related to short term goals. A detailed discussion on the inclusion of service quality variables in efficiency assessments will be presented in section 4.3 of this Chapter.

In production efficiency assessments inputs and outputs are preferably measured in physical units rather than in value. Nevertheless, as can be seen in Table 4.2, there is a large number of variables that can be used to measure production efficiency, with both quantity data and value data playing a role in these assessments.

Concerning intermediation efficiency assessments, the variables used relate, in general, to cost sources on the input side and to revenue sources on the output side. Costs that have been considered on the input side relate mostly to personnel expenses, administrative expenses, interest costs, and non-interest costs. In parametric studies, such as Zardkoohi and Kolari (1994) and Berger et al. (1997), total costs are usually used as the dependent variable of the model. In these studies the price of capital and labour are also used as inputs of the efficiency assessment.

The revenue sources used on the output side of intermediation efficiency assessments are in general interest income and non-interest income (e.g. Oral and Yolalan, 1990; Oral et al., 1992). In the study of Athanassopoulos (1997) non-interest income was used, but instead of interest income the authors used the value of the main accounts from which this income was generated (namely the value of loans, time deposit accounts, saving deposit accounts, and the value of current deposit accounts). Berger et al. (1997) also used the value of different accounts on the output side of their intermediation efficiency assessment. Note that the studies of Athanassopoulos (1997) and Berger et al. (1997) are the only ones in the bank branch literature to calculate both production and intermediation efficiency measures.

The above specification of inputs and outputs shows that intermediation efficiency is closely related with profit. According to Athanassopoulos (1997, p. 304) "the intermediation efficiency is an appropriate measure of branch profitability which gives insights on the overall cost profile of the branch per dollar of loans sold and deposits bought". Berger et al. (1994, p. 16) also acknowledge the link between profitability and the intermediation approach, referring that "the total operating plus interest cost per dollar of deposits used in the intermediation approach is an excellent indicator of the profitability per dollar raised by the branch". The link between efficiency and profit will be further detailed afterwards.

4.2.2 Models used and their Orientation: Acknowledging the new Role of Bank Branches

Most of the existing bank branch efficiency studies calculate technical efficiency rather than cost, revenue, or profit efficiency. This is especially true in non-parametric studies, since in parametric studies the use of cost functions makes the resulting efficiency measure (when it is calculated) to be economic (cost efficiency or X-efficiency[2]).

In the non-parametric DEA literature technical efficiency of bank branches has been calculated mostly by the CCR model (see Table 4.1). Some studies use the CCR model together with the BCC model so that scale efficiency measures and returns to scale can also be assessed. When price information is available other efficiency measures can be computed,

[2]X-efficiency "describes all technical and allocative efficiencies of individual firms that are not scale/scope dependent" (Frei et al., 2000).

namely overall economic efficiency and allocative efficiency. Only two studies in the DEA literature have used price information to calculate this type of efficiencies, namely Schaffnit et al. (1997) and Hartman et al. (2001). Both studies analysed cost efficiency, but the approaches followed were different. Hartman et al. (2001) used actual price information in their models and so calculations of cost and allocative efficiency were performed as explained in section 2.3.1 of Chapter 2. Schaffnit et al. (1997), on the other hand, used ARs to arrive at some notion of cost efficiency as explained in section 2.3.5 of Chapter 2.

Most of the studies on branch efficiency (see Table 4.1) adopt a cost minimisation perspective[3]. Under this perspective bank branches are assumed to be in control of inputs, which should be minimised. A cost minimisation perspective is perfectly understandable when banks as a whole are being analysed. However, when the unit of analysis is the bank branch this cost minimisation perspective might not be reasonable especially if most of the bank branch's costs are not controllable by branches' managers. Controllable operational costs of branches are mostly related to staff, equipment and supplies. However, "more often than not, decisions regarding human resources, location, technology, etc., are not made at the branch level, as branch networks are managed centrally" (Athanassopoulos et al., 2000, p. 360). This means that by using such factors as inputs and choosing an input minimisation perspective the level of management one is assessing does not relate to the branch but to a higher level of management. Even when decisions relating to staff are undertaken at the branch level there is not discretion regarding this input. For example, in the case of the Portuguese branches, which are the subject of our empirical study, branch managers can decide whether they need more or less staff, but the ultimate decision concerning the hiring/firing of staff is made at a higher level of management.

At the same time cost cuts are mainly a concern of higher levels of management rather than branch management. A bank manager said to us that "the best way to cut branches' costs was to close them", and for that reason branches should focus on what they have been created for - selling, serving and capturing clients. While these arguments support the choice of output oriented measures, the literature does not match this reasoning. The argument that is most commonly pointed out for using input oriented measures is the need to rationalise resources at existing branches (e.g. Camanho and Dyson, 1999). In the work of Sherman and Ladino (1995) the authors argued that an input orientation

[3]See also appendix 1 of Camanho and Dyson (1999) where from the 15 studies reported 13 used input oriented models.

was preferred as the bank under analysis wanted to expand by increasing the number of branches, financing part of the cost through savings in the existing branch network. According to Sherman and Ladino (1995) banks can manage less easily the expansion of service volume. Schaffnit et al. (1997, p. 278) corroborate this argument as "branches, in general, have no direct control over the amount of services their customers require". While this is true for typical transactions at the branch (withdrawals, deposits, general inquiries, cheques cashed, money transfers, etc.) it is questionable for new sales and attraction of new clients where branch personnel might, and should, have and active role. Indeed, the changing role of bank branches towards selling and marketing oriented activities is not a passive one, but a proactive role that demands efforts at the branch for increasing the sales of new products and capturing new clients. Input oriented measures seem, therefore, to be consistent with the traditional transactional role of bank branches, but its new role as marketing and retailing centers seems to demand for output oriented measures.

Some studies in the literature have used output oriented measures but few have justified this orientation based on the new role of bank branches towards selling rather than towards servicing. In most cases the choice for output oriented measures or non-oriented measures was justified by atypical measures of efficiency that were being analysed [like quality efficiency in Soteriou and Stavrinides (1997); Soteriou and Zenios (1999) and market efficiency in Athanassopoulos (1998)], by different objectives of analysis [like measuring the success in achieving the established targets in Lovell and Pastor (1997)], or by some theoretical considerations [like in Haag and Jaska (1995)].

Golany and Storbeck (1999) used an output orientation in the assessment of production efficiency, in line with a view about the role of bank branches towards selling. In fact, the argument of Golany and Storbeck (1999) for using an output oriented measure was that the bank was "in the midst of a growth period" and as such focusing on output enhancements rather than on input savings. Avkiran (1999a, p. 211) used a similar argument for choosing an output oriented DEA model mentioning that "during an exercise to expand market share of banking products, the focus could shift to output maximisation". Kantor and Maital (1999) also used output oriented models to measure the efficiency of two activities of bank branches (providing service to customers, and carrying out transactions), but the authors do not mention the reasons for choosing such an orientation.

Interestingly most recent studies that put forward explicitly their concern of accounting for the changing role of bank branches, like those of Cook et al. (2000) and Cook and

Hababou (2001), do not use output oriented measures. In Cook et al. (2000) the authors distinguish between two types of activities that happen in bank branches: transactions (or service) and sales, and calculate an overall efficiency measure that is an aggregate of sales and service efficiency measures. The inputs of the model were staff related (service staff, sales staff, support staff and other staff) and the outputs were related with service (counter level deposits, and transfers between accounts) and with sales (retirement savings plan openings, and mortgage accounts opened). The authors used, however, an input oriented model, which does not seem appropriate for the purpose of increasing sales - the fundamental objective of bank branches when one assumes, as the authors did, that they are evolving from a traditional role to a "more general and proactive function as universal financial agents with a distinct *sales* culture" (Cook et al., 2000, p. 209). In a continuation of the above study, Cook and Hababou (2001) also consider a set of inputs and outputs both related to sales and service and calculate a service and sales measure of efficiency through a goal programming model. It is worth noting that the reason for Cook et al. (2000) and Cook and Hababou (2001) not using output oriented models might be related with their objectives of analysis. Indeed, these authors were not so much concerned with establishing targets but more with comparing the performance of bank branches on the two dimensions considered critical: service and sales.

It is important to note that the use of output oriented measures requires the use of market information on the input side (Athanassopoulos et al., 2000), as the ability of a bank branch to increase its sales and customers depends on the market potential of the location where the branch is in. This is an additional difficulty in using this direction of improvement as market information is not easily gathered. In the study of Golany and Storbeck (1999) the authors attempted to include a set of non-discretionary factors related to market size, economic status of the area, and competitive activity as these "are factors that influence the activity of the branch but are not under its direct control" (Golany and Storbeck, 1999, pp. 17-18). However, most of the factors analysed by the authors were not incorporated in the final data set used for lack of quality or evidence that they affected the performance of branches (the only environmental factor used was a measure of unemployment).

There are also studies in the literature that when acknowledging the new role of bank branches focus the efficiency analysis on completely different dimensions from those traditionally studied. These are the cases of Drake and Howcroft (1995)[4] that analysed the selling function of bank branches, and Nash and Karwat-Sterna (1996) that analysed the cross-selling efficiency of bank branches. The analysis of the selling function in Drake and Howcroft (1995) led to substantial differences from traditional studies. For example, the outputs of the DEA model were mainly sales related (sales of personal loans, cheque accounts, mortgage loans, etc.), while the choice of inputs was guided by the need to consider the "existing customer base in each of the most important balance sheet categories which forms the foundation for cross-selling and repeat business opportunities" (Drake and Howcroft, 1995, p. 12). It is worth noting the inclusion of the design or layout of the branch in the efficiency assessment of Drake and Howcroft (1995), and the corresponding consideration that new layouts are required so that the change in the role of bank branches from 'transactional' to sales oriented becomes effective.

The study of Nash and Karwat-Sterna (1996) also differs substantially from traditional bank branch efficiency studies. The authors considered the selling of four related products in their analysis of the cross-selling efficiency of 75 branches. The DEA model used was the additive model and only outputs were used. These outputs assumed the form of ratios [number of products sold together with housing loans (main product in analysis) divided by the number of housing loans approvals].

4.3 Efficiency and Service Quality

In Chapter 3 we mentioned the importance of considering service quality in efficiency assessments of service organisations. When the objective of the analysis is to measure efficiency, quality issues cannot remain ignored especially if we agree that providing better quality is likely to consume extra resources.

Banking efficiency studies that include the quality of the outputs produced have recently appeared in the literature concerning the assessment of bank branches. Studies on efficiency assessments of banks usually rely on accounting published data that do not include service quality variables. This is maybe the reason why the concerns about quality in banking have first arisen in the bank branch efficiency literature and not on the efficiency literature concerning banks[5].

[4]This study is not included in Table 4.1 because it is a working paper. In order to be exhaustive we decided to include in Table 4.1 only published papers.

[5]The only exceptions are, to the the best of our knowledge, the work of Faulhaber (1995) who included variables related to customer satisfaction on a bank efficiency assessment, and the work of Mukherjee et al.

The inclusion of service quality variables in DEA assessments is not well established yet and no consensus exists concerning how this variable relates with efficiency. One can find in the literature studies that include service quality in a *post-hoc* analysis and studies that include service quality variables in the actual technical efficiency measurement. In the latter case the doubt concerning whether service quality is an input or an output arises, and studies can be found where service quality variables are either inputs, outputs or both. The main assumption behind these different treatments is whether service quality is a function of efficiency [SQ = f(Eff)] or the reverse [Eff = f(SQ)]. In the first case the type of analysis that makes sense is a *post-hoc* analysis, while in the second case the service quality should be considered an input or output of the efficiency measure. Another line of research is the one that uses service quality variables to actually measure service quality through DEA (e.g. Soteriou and Stavrinides, 1997, 2000; Soteriou and Zenios, 1999; Manandhar and Tang, 2001). This is not the kind of studies we are interested in, since service quality will be considered as given and not a factor to be measured. These studies may, however, provide interesting insights on the linkage between efficiency and quality.

The existing studies in the literature that include service quality dimensions in the efficiency assessment are not many. The first was the study by Parkan (1987) where quality variables were introduced both on the input and output side of a CCR input model aimed to measure production efficiency. The quality variables introduced on the input side were a ranking of the quality of customer service area, and a ranking of the marketing activity of the branch[6]. On the output side Parkan (1987) used a customer service survey ranking and the number of corrections per number of transactions which is an indicator of the quality of the process of providing the service[7]. Avkiran (1999a) included a service quality variable on the input side of a production efficiency assessment. The variable considered was 'staff conduct', which was the dominant factor in a service quality framework developed in Avkiran (1994, 1999b). This variable is defined as "responsiveness, civilised conduct and presentation of branch staff that will project a professional image to the customers" (Avkiran, 1999a, p. 210). Golany and Storbeck (1999) included two outputs in their production efficiency assessment that are related with service quality. One of these outputs was depth of relations (measured by the number of accounts per customer) intending to capture the

(2003) who measured quality efficiency of Indian public sector banks.

[6]Parkan (1987) used the inverse of the marketing activity ranking, as less of this input would mean more marketing activity.

[7]This output was also inverted, as less of it is preferable.

customers' loyalty to the branch, and the other was a customer satisfaction index based on "a quarterly survey the bank runs" (Golany and Storbeck, 1999, p. 19). Another study where quality variables are included as part of the technical efficiency assessment is that of Athanassopoulos (2000). The author used a two stage DEA model to assess service quality and technical efficiency, applying the capabilities-service quality-performance (C-SQ-P) triad of Roth and Jackson III (1995). In a first stage service quality is assessed through DEA, and in a second stage the quality targets resulting from the first stage (on corporate, interactive and physical quality) are used as inputs of the technical efficiency measurement. Athanassopoulos (2000) explicitly puts forward the following relationship between efficiency and service quality: Performance = f(Service quality, Branch size, Cost, Outputs, Work load), understanding thus service quality as a cause of performance and not an effect [i.e. Eff = f(SQ)]. This view is shared by some other authors like Roth and Jackson III (1995) who hypothesised that performance is directly affected by service quality. The authors tested this hypothesis on data gathered from a survey on retail banking and found supportive evidence for this link.

In summary only four DEA studies incorporated service quality variables in the measurement of production efficiency. One uses service quality variables in both sides of the assessment (Parkan, 1987), two studies consider service quality variables as inputs of the efficiency assessment (Avkiran, 1999a; Athanassopoulos, 2000), and one study considers these variables as outputs of the assessment (Golany and Storbeck, 1999).

The studies that include service quality variables in a *post-hoc* analysis are two. The first is the study of Schaffnit et al. (1997), which considers that service quality, among other factors, is neither an input nor an output of the efficiency assessment. In the analysis undertaken the authors divided the factors to be analysed in effects and causes of performance. Quality was classified in the effects side, that is as a factor that is potentially affected by performance [SQ = f(Eff)]. The results from Schaffnit et al. (1997) indicate a positive effect of efficiency scores on the quality variable used (a customer service index). The second study that uses service quality variables in a *post-hoc* analysis is that of Athanassopoulos (1997). In this study target outputs resulting from the intermediation and productive efficiency assessments were regressed on service variables obtained from a questionnaire administered to branch customers. In the system of regression equations the dependent variables used by the author were the various output targets, and the independent variables were ten quality factors and two inputs (full-time equivalents and number of computer terminals). Clearly this author considers that [Eff = f(SQ)] as in both his studies (Athanassopoulos, 1997, 2000) this relationship is put forward, although treated differently. From the analysis undertaken the author could provide insights regarding the effects of each quality variable on outputs, and regarding the influence of service quality on the efficiency of each branch. Conclusions point to a positive but not statistically significant impact of physical components of service quality on output targets. On the other hand, network size and image of the branch had a significant positive effect on branches' outputs, and service time had a negative and significant effect in almost all branches outputs.

As noted earlier, some studies follow a different approach to incorporating quality dimensions into the analysis by actually measuring the service quality of bank branches through DEA. For example, Soteriou and Stavrinides (1997) presented an indicator of *service quality efficiency* defined as how well the branch transforms resources (personnel, space, computers, etc), given a certain account structure of their clientele, into service quality (the only output of the assessment). Soteriou and Zenios (1999) presented a DEA model based on the (C-SQ-P) triad (see Roth and Jackson III, 1995) and on the service profit chain of Heskett et al. (1994). Based on these two frameworks, and making some adaptations, Soteriou and Zenios (1999) link operations, service quality, and profitability performing three efficiency assessments. The set of inputs used in each of these assessments was very similar but the output changed with the type of efficiency being assessed. In order to measure operational efficiency they used the total time involved in processing the tasks for all transactions carried out at the branch. The output of the service quality efficiency model was a measure of service quality obtained through questionnaires administered to internal customers[8], and the output of the profitability efficiency model was the profit generated at each branch. Results from these models pointed to a strong and positive relation between service quality and operational efficiency. However, concerning service quality and profitability results showed no relationship between short-term profitability and service quality. Manandhar and Tang (2001) developed a framework for assessing bank branches efficiency that is very similar to the one of Soteriou and Zenios (1999). The authors propose the use of the model of Lovell and Pastor (1997) to aggregate efficiency measures resulting from each of the three dimensions analysed (internal service quality efficiency, operating efficiency, and profitability efficiency). In this model efficiency measures resulting from each of the three assessments are regarded as outputs of a DEA model that has no inputs. Manandhar and

[8]The questionnaire was based on the framework proposed by Parasuraman et al. (1988) - SERVQUAL.

Tang (2001) put forward this framework but do not apply it in any empirical application (this is the reason why this study is not included in Table 4.1).

Our concerns around the service quality dimension are particularly related with the importance of considering service quality variables in traditional production (or intermediation) efficiency assessments rather than with measuring service quality efficiency through DEA. In fact, we raise some doubts concerning the practical utility of service quality DEA models, whose single output is a service quality index, namely on their ability to provide suggestions for quality improvements. The resulting efficiency measures might be of some use in the sense that they indicate the extent to which the maximum potential on SQ is being attained given a set of resources used. But even in this case the practical consequences of such efficiency measurement are in our opinion dubious. The suggestions one could give to managers in face of such efficiency measure are either that they should increase their SQ (what they already know because the ideal objective should be to attain the maximum possible on this index, which would be synonymous with excellence), or that they should decrease some inputs in some amounts (which is likely to improve technical efficiency but not necessarily improve service quality because customer's perceptions on service quality could change giving an altered input mix (Grönroos, 2000)).

Our understanding of service quality and its relationship with efficiency points to the consideration that service quality is a component of efficiency when service organisations are considered (Grönroos, 2000)). This means that even if no quality variables are accounted for in the analysis, quality is being implicity considered. In fact, a high efficiency measure necessarily indicates that a service firm is producing good quality of outputs, as under the reverse hypothesis the firm would lose customers in the long run, and as such efficiency would deteriorate (recall that in service organisations outputs relate to sales and not to products that were produced and not yet sold). This does not mean that service quality variables should not be explicitly included in the efficiency measurement. If available they may in fact add value to production or intermediation efficiency assessments, as a bank or bank branch uses its resources to produce a given amount of services that are not only characterised by their quantity but also by their quality.

4.4 Efficiency and Profitability

As mentioned in Chapter 3, a profit analysis may be a way of incorporating a quality dimension in efficiency assessments. In addition, banks and bank branches are for-profit organisations and therefore it is important to measure the extent to which the profit objective is being attained (Athanassopoulos and Thanassoulis, 1995).

Some studies analysing the efficiency of bank branches (referred to in Table 4.1) linked efficiency and profitability. The way these two concepts were linked followed essentially two routes. One route consists in analysing the link between efficiency and some profitability measure *a posteriori*. This was the route followed by Schaffnit et al. (1997) who regressed a profit measure (net profit per million dollars of total retail business volume) on the efficiency scores. Conclusions pointed to a statistically significant positive effect of performance on profits. Camanho and Dyson (1999) also analysed the link between efficiency and profitability through an analysis *a posteriori* using an "efficiency-profitability matrix". The profit measure used was the ratio of net profit before indirect costs and total costs. The general trend showed that high efficiency was associated with high profitability.

In the above mentioned studies the profit dimension is included through the use of some accounting measure of profit. An alternative route is to measure profit efficiency concerning what could maximum profit have been controlling for various factors. This has been done through what has been called by some authors profitability efficiency, where a set of expenses are considered on the input side and a set of revenues are considered on the output side. Such an approach is similar to the intermediation approach to bank branch efficiency measurement. Oral and Yolalan (1990), who were the first to analyse the relationship between efficiency and profits, followed this route. The authors calculated profitability through a DEA model where inputs were a set of expenses (personnel, administrative, depreciation and interest) and outputs were interest earned on loans and non-interest income. Results have shown that profitability and efficiency were closely related (see also Oral et al. (1992), which is an extension of Oral and Yolalan (1990)). Another study using this approach was that of Athanassopoulos (1997).

The main issue arising when the intermediation approach is used for profit efficiency measurement is the type of model orientation that should be used. In order to maximise profits bank branches seek at the same time to minimise costs and to increase revenues. In this sense a non-oriented efficiency model seems to be the ideal means of accounting for this two-way intention. Such an approach was the one used by Athanassopoulos (1997), but Oral and Yolalan (1990) and Oral et al. (1992) used input oriented measures, meaning that their model was mostly a cost model rather than a profit model.

Profitability efficiency has also been calculated in the literature following other means.

Namely Soteriou and Zenios (1999) used a DEA model to assess profitability efficiency, where the only output considered was profit. The authors also calculated operational efficiency and represented results on a "efficiency-profitability matrix" but did not conclude about the relationship between profits and efficiency. An analysis of this matrix in Soteriou and Zenios (1999, p. 1234) shows that most branches are in the "Dogs"[9] quadrant. Nevertheless, this matrix does not seem conclusive as very few branches seem to be represented in there (and not the full sample of 143 bank branches).

Although the two routes mentioned above seem valid for linking efficiency and profit they do not account for a real profit efficiency measure as referred to by Färe and Primont (1995) where factor's prices are included in the analysis, and where profit is sought to be maximised. In Chapter 8 we review the literature on existing profit efficiency measures and propose a new way to measure profit efficiency through non-parametric methodologies. Note that in the parametric banks' efficiency literature the measurement of profit efficiency is now quite usual after the work of Berger et al. (1993) fostering the use of profit functions as a way to account for quality differences between banks (see also Berger and Humphrey, 1997; De Young and Nolle, 1996; Berger and Mester, 2000). According to Berger et al. (1993) the use of profit functions brings in other advantages apart from those relating to the inclusion of a quality dimension. For example, a profit function allows a better understanding of the sources of inefficiency, namely whether these are more related to inputs (costs) or to outputs (revenues). Interestingly, there is not in the literature any study measuring the efficiency of bank branches through profit functions.

4.5 Summary

In this Chapter we reviewed the literature on bank branches efficiency assessment. In this review we focused particularly on the aspects that are crucial to the development of our own model to assess bank branches efficiency, which is presented in the next Chapter. In this sense we start by enumerating the various bank branches' efficiency studies in the literature but analyse them only in two respects: (i) the type of data and the type of inputs and outputs used, (ii) and the extent to which existing studies acknowledge the new role of bank branches. An analysis of the type of inputs and outputs that have been most used in the literature is important as a means to justify our own input/output choice. In addition,

[9]Branches experiencing high operating efficiency and low profits.

as it is our intention to acknowledge the new role of bank branches in our empirical analysis, it is important to analyse the extent to which previous studies have done so, and how they have done so.

In this Chapter we also analyse the way service quality and profitability have been incorporated into efficiency analysis. The review of the literature concerning this aspect will be important in the development of our own model to assess bank branches efficiency since we intend to account for service quality and profitability in our efficiency assessments. Being bank branches service and for-profit organisations these two aspects are of utmost importance in analysing their overall performance.

Chapter 5

Measuring the Performance of Portuguese Bank Branches

From what has been said in previous Chapters it seems clear that new forms of measuring the efficiency of bank branches are needed. Such efficiency measures should account for the new role of bank branches, taking into account the fact that these are service organisations, and also for-profit organisations. In this Chapter we present the model that is to be used in subsequent chapters for analysing the efficiency of a sample of Portuguese bank branches.

5.1 Model for Bank Branch Efficiency Assessment

The previous chapters reviewed the triad between service quality, efficiency, and profits. This triad represents a starting point in our development of a model to analyse bank branches efficiency. In developing this model we attended not only to this triad but also to the changing role of bank branches from transactional based to sales based. At the same time the opinions and concerns of the managers of the bank under analysis were also of utmost importance in establishing our own model.

According to some opinions of Portuguese managers of the bank under analysis the growing use of new distribution channels does not constitute a threat to the survival of bank branches, at least in the short run. This fact is, instead, understood as an opportunity since bank branches can move a number of costly activities from branches to these new channels. Traditional transactional activities (like withdrawals, deposits, general enquiries on banks products, money transfers, etc.) consume human resource's time that is more expensive than ATM's time or internet and phone facilities' time. If this type of activities moves from branches to other channels then there is scope for increasing efficiency and profitability of bank branches as long as customers do not see these changes as a reduction in the quality of the service provided. Seeing new means of distribution as an opportunity for increasing profits, banks, and particularly the bank under analysis, have attributed to its branches the responsibility for motivating customers to use other distribution channels. This means that the bank under analysis recognises that the role of bank branches is changing from transactional based to sales based. In fact transactions are to be moved as much as possible to alternative and cheaper distribution means, while personnel at the branch shall be concerned mostly with selling, cross-selling, and managing the mix of accounts of their clients such that the rentability of each customer for the bank as a whole is maximised.

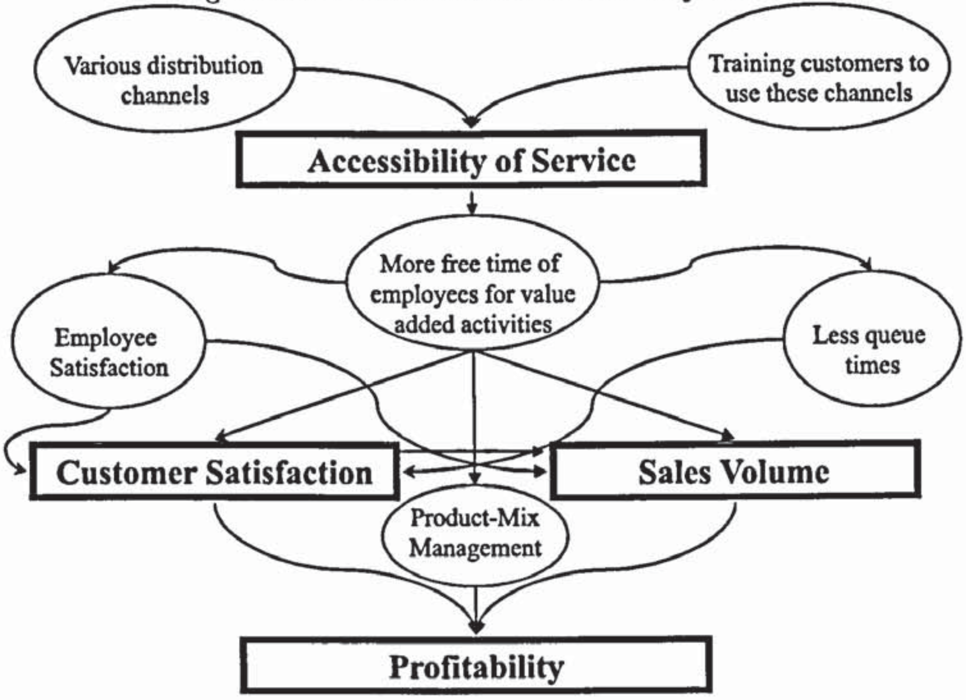
In order to measure efficiency of bank branches it is important not only to be aware of the new role of bank branches, but also to be aware of the main objectives of bank branches. In general, and based on the empirical work performed, we can say that bank branches have the following objectives:

- ♣ To allow their customers to access banking services in any place in the country (or abroad), which is done by providing several different distribution channels of their services (including internet, phone lines, and ATM machines);
- ♣ To foster the use of new distribution channels so that branch personnel can use their time on value-added activities;
- ♣ To increase sales and the customer base of the branch, while serving the clients that visit the branch with a high level of service quality;
- ♣ To manage the product mix in a way that generates high profitability, without reducing service quality associated with any product;
- ♣ To have the maximum possible profitability given bank branches resources and environment.

The first two objectives are related to the accessibility of banking services, the third with sales volume and operational capabilities, and the last two objectives are related to product mix management and profitability. The efficiency of a bank branch can, therefore, be defined as the extent to which these objectives are attained given the external environment the branch is in.

The objectives of bank branches stated above are naturally inter-related. The inter-relations between these can be seen in Figure 5.1, which served as a basis for the development of some efficiency measures meaningful in the new context of bank branches.

Figure 5.1: Model of Bank Branches Objectives



The model in Figure 5.1 links different bank branch objectives. To provide a good accessibility of service is one of the main objectives of banks. This is attained by providing customers with a large number of available service distribution options. However, this might not be enough if customers do not know how to use the various distribution channels. For that reason, an effective accessibility of banking services is a result not only of the existence of a variety of distribution means, but also of training and motivating customers on using these means[1]. This is a responsibility of branches which are the main interface between the bank and the customer. The accessibility of service is considered by some authors as a dimension of service quality (e.g. Parasuraman et al., 1985). Although the type of accessibility that is dealt with in the literature is not directly related with the existence of alternative distribution channels[2], we believe this is a factor that contributes to increase

[1]The motivation for increasing the use of other distribution channels passes also through pricing policies.

[2]For example, Parasuraman et al. (1985) mention that access refers to approachability and ease of contact (Parasuraman et al., 1985, p. 47).

service quality as perceived by the customer.

When a good accessibility of service is attained, branch's personnel will devote less time to transaction activities and will have more time for performing other activities, such as selling new products, attracting new customers, managing the mix of products, etc. This will have a direct impact on employee satisfaction (they will be concerned with more interesting activities), and on queue times (as less customers go to the branch for transaction activities). The lower pressure on employees generated by less queues and their higher motivation will have an impact on their willingness to respond to customer queries and to inform customers. These factors, more satisfied employees, smaller queues, and higher service quality due to higher accessibility of banking services, are likely to improve customer satisfaction[3]. Higher customer satisfaction will have impacts on customer retention (see Zeithaml et al., 1996, for details on the relationship between service quality and customer behavioural intentions of remaining with or defecting from the company), which increases sales volume and market share (Rust et al., 1995). Favourable word of mouth, spread by satisfied customers, has effects both on retaining existing customers and attracting new ones (Kordupleski et al., 1993). These factors, will positively affect market share and sales volume. Sales volume is also likely to be influenced by the fact that employees will devote their time to sell new products to existing customers or use more pro-active strategies to attract new ones. At the same time new technologies allow banks to centralise information on customers' preferences and behaviour. This enables banks to create specific products to satisfy customers' needs, which in theory increases customer loyalty and as a result increases sales volume.

If profit is to be maximised then product mix management should be emphasised. In the opinion of some bank managers interviewed, it is important that branches move clients from less profitable products, such as time deposits, to more profitable ones such as funds, or similar products. For doing this, branch staff need more free time, which increases as the usage of the alternative distribution channels increases. The product mix management and the increased sales volume will both contribute to increase profits. Higher profits might also be a result of superior service quality levels, since high levels of service quality result usually in reduced price elasticities. That is, satisfied and loyal customers are willing to pay

[3]The positive relationship between employee satisfaction and customer satisfaction is very well documented in Schneider (1991). The relationship of these variables with profit is also explored by the author. Heskett et al. (1994) also provides a theoretical framework according to which employees satisfaction plays a role in determining customers satisfaction.

for the benefits they receive and are eventually more tolerant to price increases (Anderson et al., 1994). In addition, the reputation of the firm in the market, resulting from its high levels of service quality, is also likely to allow the company to charge a price premium and thus increasing profits (Zeithaml, 2000).

Note that relationships between service quality and profitability have been extensively analysed in the Marketing literature namely in Zeithaml (2000) who present an extensive literature review on this subject, and Rust et al. (1995), who provide a conceptual framework linking service quality improvement efforts to profitability. In this literature the study of Loveman (1998) is of particular relevance since the author analysed 450 bank branches. Results from this study suggest that there is a relationship between internal service quality (rewards, quality of bank management, etc) and employee satisfaction. There was also supporting evidence suggesting that employee satisfaction is related to loyalty of employees, and that customer satisfaction is related to customer loyalty.

On purpose we have excluded from the model in Figure 5.1 certain variables. These are costs, which impact directly on profit, and customer loyalty. Concerning costs the Marketing literature in general believes that higher service quality leads to reduced costs (see e.g. Rust et al. (1995)). The effect of service quality on costs is mainly due to the retention of customers, since obtaining a new customer costs 5 times more than keeping an existing customer (Zeithaml, 2000). Indeed, a new customer implies a set of activities (like adding the customer to the database, giving detailed information on what will be expected from the service, etc.) which are expensive, whereas long term customers know what to expect and ask fewer questions (Reichheld and Sasser, 1990). Nevertheless, the impact on banking costs does not come from a single source, and in the present banking context there are no certainties that increased accessibility of service may lead to cost reductions. There are strong expectations of transactional cost reductions due to the introduction of remote banking (see ECB, 1999a). Nevertheless, concerning overall costs there are uncertainties on the effects of IT investments. According to the ECB (1999a) some of the reasons why overall costs are not likely to decrease, at least in the short run, relate with high initial investments and maintenance costs, and with the fact that some time is needed for customers to change their habits. This forces banks to maintain duplicate capacities for different customer segments. ECB (1999a) reports that bank costs are only expected to decrease in the long run, and this reduction is not expected to be dramatic. Staff costs, which account for a high percentage of overall costs, are not expected to decrease strongly due to the need of more qualified personnel in other activities like marketing and IT.

The loyalty issue preoccupies bank managers of the bank under analysis, as they expressed their concerns with the effects of remote banking on the loyalty of customers. Indeed, with IT developments banks can be more informed about customers, but the reverse is also true, i.e. customers can have easy access to banks information and comparisons are easy to make. At the same time customers do not visit the branch so often, and as such the personal relationship between branch employees and customers will slowly vanish. According to ECB (1999a) one cannot yet observe a significant decrease in customer loyalty. "The importance of personal contact is expected to decline only gradually. However, there are increasing signs that customers have become more mobile in their search for the best offer in the market" (see ECB, 1999a, p. 21).

5.2 Efficiency Measures

Given the three main objectives of bank branches (accessibility, sales volume, and profitability) discussed in the previous section, and the model that has been constructed linking these objectives, we developed three efficiency measures reflecting the extent to which these objectives are being accomplished by bank branches. The first efficiency measure is called **Transactional Efficiency** and is intended to measure the extent to which, given accessibility of services, customers use new distribution channels for transactional activities. The second efficiency measure is called **Operational Efficiency** and is intended to measure the extent to which a branch, given its resources and environment, increases its sales and its customer base, while serving its existing customers with adequate service quality. The third measure of efficiency is called **Profit Efficiency** and is intended to measure the extend to which a branch is maximising profit.

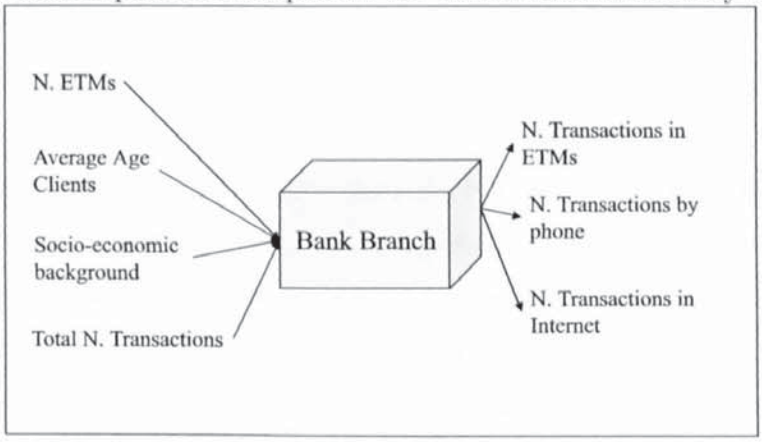
Each of these efficiency measures is detailed in the next sections, where we present the set of inputs and outputs that should ideally be used, and the set of inputs and outputs that will in fact be used given data availability.

5.2.1 Transactional Efficiency

This measure of efficiency relates to the objective of bank branches to move from transactional activities to value-added activities. It is important to stress at this point that we understand transactions as being all the types of activity performed by branch personnel that take place at the bank branch but could have been done through other means of distribution. In this sense, deposits, withdrawals, balance enquiries, request of cheque books and so on are understood as transactions.

Ideal inputs and outputs of the transactional efficiency assessment can be seen in Figure 5.2.

Figure 5.2: Ideal Inputs and Outputs in the Transactional Efficiency Assessment



Inputs to be used relate to the relevant resources at the branch for its transactional activities and to environmental aspects that impact on the ability of customers to use alternative distribution means. The resources at the branch are the number of ETM's (electronic teller machines), which include ATM's and also other equipment such as cheque dispenser machines (CATs). One of the environmental factors is the location of the bank branch and the socio-economic background of its clients that may impact on their willingness to use alternative distribution channels. For example, in rural areas a high effort of the branch personnel to motivate clients to use the internet might be in vain because clients may not have the necessary knowledge, equipment, and openness to use this distribution channel. The age of clients may also influence their propensity to use new distribution means. Younger clients have higher openness to try new things and are likely to access the internet more easily and to use ETMs more often than older and more traditional clients. This variable might be, however, difficult to collect.

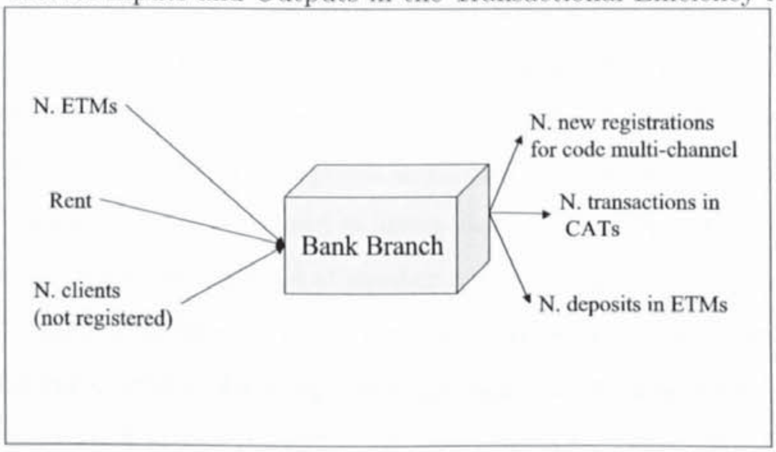
The input 'total number of transactions' includes transactions performed at the bank branch by branch's staff and also transactions performed in alternative distribution channels. This input is considered because we want to evaluate the extent to which the branch moves transactions to new distribution channels. If no other inputs were considered, one would have an efficiency measure given by the ratio between the number of transactions in each alternative distribution channel (the outputs) and the total number of transactions.

A ratio of 100% would mean that all the transactions were done in alternative distribution channels and not by the branch's personnel at the bank branch. This would mean that our objective was completely accomplished.

The transactional measure of efficiency is naturally oriented towards the maximisation of outputs. The ideal outputs pointed out for this assessment are self explanatory. The only aspect worth noting is the fact that in principle outputs should account for transactions done in the various means by branch clients only (since this data are potentially easier to collect). Obviously, the branch may put some effort in motivating clients that are not their own in using alternative distribution channels, but considering these customers would result in unnecessary complications to the analysis. This clearly represents a limitation of the assessment, as some branches located in high passing trade zones may have more transactions performed by non-branch clients than by their own clients. This fact may under-estimate the efforts of a branch in motivating customers (irrespective of the branch in which they are clients) to use alternative distribution channels.

The foregoing inputs and outputs would be ideal for measuring transactional efficiency. Unfortunately most of the data in Figure 5.2 are very difficult to collect, and this was the case in our study. The data on transactions that we have available allow for the measurement of transactional efficiency through the set of inputs and outputs shown in Figure 5.3.

Figure 5.3: Actual Inputs and Outputs in the Transactional Efficiency Assessment



The type of inputs that we consider are similar to the ideal inputs except that we use some surrogates for variables on which we have no data. The socio-economic background of the area where the branch is located could not be provided by the bank, and we used 'Rent' as a surrogate for this variable. Rent is a variable used regularly by the bank for proxying both the location and the size of the bank branch. Concerning the ideal input 'total number the transactions' we could not use this input because the data on transactions that the bank collects includes only those performed at the branch by the branch's personnel and those performed in ETMs located in the bank branch, where both branch's clients and non-clients are considered. On the input side we also use the number of clients not yet registered for multi-channel use, which links with the output number of registrations for a multi-channel code (the use of this input will be further explained afterwards).

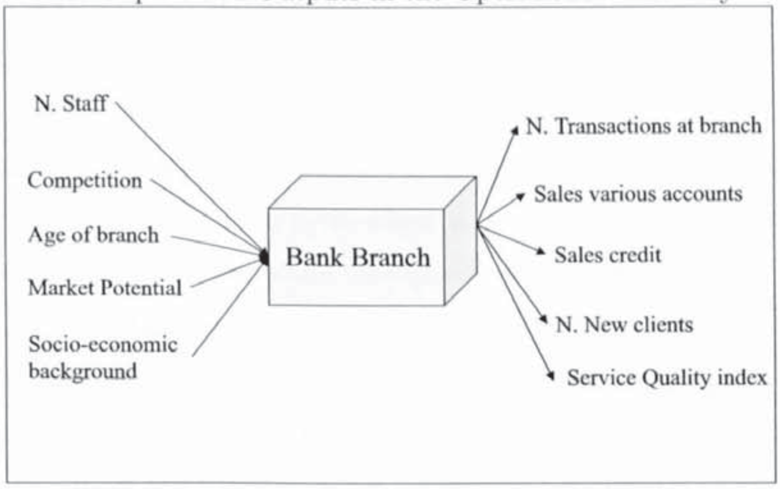
On the output side we could not consider any of the ideal outputs. Indeed, the information that the bank has available concerns the total number of transactions performed in ATMs and CATs located in the bank branch, and these may include clients and non-clients of the bank branch. For the case of transactions in ATMs these may, in fact, relate to a customer from any bank in the country or abroad and therefore the number of transactions performed in ATMs should not be used in our assessments. There is, however, one type of transaction that is mostly undertaken by customer's of the branch: deposits. A deposit in an ATM or CAT of a given branch was necessarily performed by a client of the bank as a whole, and it is likely that the deposit was done by a client of the branch where the equipment is located. In Portugal ATMs are provided to all banks by an entity called SIBS. CATs are, however, equipment that belongs to the bank, and is to be used only by clients of the bank (note that some ATMs may also be property of the bank but usually they are not). This means that transactions performed in CATs are also likely to have been done by clients of the bank branch where the CAT is located and therefore we can use this variable in our transactional efficiency assessments. We have, therefore, two output variables indicating the degree of usage of ETMs in the bank branch (number of transactions in CATs and number of deposits in ETMs), and also a variable indicating how many branch's customers have registered to access internet in each month (all our data are monthly). This variable is used instead of number of transactions in internet since the bank could not supply the actual number of transactions performed in the internet by branch's clients. The number of clients that registered in each month links with the input N. of clients not yet registered at the beginning of each month for using this channel. It was important to consider this input rather than number of clients, since we are interested in assessing how effective a branch is in motivating those customers that do not yet access the internet to use this new distribution channel.

5.2.2 Operational Efficiency

The operational measure of efficiency is related to all types of operations that go on in a bank branch. It accounts especially for value-added operations (sales related) and therefore this measure is linked with the previously defined transactional efficiency. That is, the more effective a bank branch is in moving transactions to alternative distribution channels the more time branch staff will have to perform value-added activities, and therefore it is likely that the operational efficiency is higher for those bank branches which show higher transactional efficiency. This link is analysed in detail, and confirmed, in Chapter 11.

Ideal inputs and outputs of the operational efficiency assessment are presented in Figure 5.4.

Figure 5.4: Ideal Inputs and Outputs in the Operational Efficiency Assessment



This measure presents similarities with the usual production efficiency measures in the banking literature but it has some differences. Firstly it has something of a market efficiency measure as defined in Athanassopoulos and Thanassoulis (1995), and Athanassopoulos (1998). This is so because one of the objectives of the branches is to increase their sales and customer base. The ability of the branch to do this, depends on the market conditions it faces, like the degree of competition in the marketplace, the location of the branch, or the socio-economic conditions of people in this location (which are potential customers of the branch). Athanassopoulos (1998, p. 174) defines market efficiency as "the extent to which individual bank branches, given their capacity and resources available, utilize their market potential by maximizing sales". Such measure of efficiency is naturally output oriented. Our measure of operational efficiency fits the definition above since it considers on the output side the sales of various accounts, but it also incorporates other operational aspects. Indeed, alongside with the objective of bank branches to increase sales volume and their customers' base, the branch cannot stop serving customers that do not go there for buying a product but for performing general transactions that could have been done in alternative distribution channels. Therefore, on the one hand branch managers would like to reduce general transactions at the branch, so that sales could be improved, but on the other hand they need to serve customers whatever they require in their visit to the branch. Transactions appear, therefore, as an output of the operational efficiency assessment but it is non-discretionary in the sense that the branch does not want to increase it.

In accordance with the above we consider in Figure 5.4 inputs relating with the environmental conditions and operational conditions faced by the branch. Environmental factors relate with the level of competition, the market potential, the socio-economic background of the branch's clients and the branch's age. The socio-economic background and the market potential are intended to reflect, respectively, the 'quality' and the 'size' of the market where potential customers are in. The consideration of the age of a bank branch is important when the type of outputs used capture growth as in our case. In fact, bank branches pass usually through a life cycle where in their first years they experience high growth rates in sales and customers and then they reach a maturity phase where growth rates are lower. Under this circumstance it is expected that younger bank branches have higher growth rates than older bank branches.

The input competition may pose some problems in the assessment since it is not always clear whether more competition or less competition is better. Athanassopoulos and Thanassoulis (1995) considered competition as a negative input, and as such the inverse of number of competing establishments in the surrounding area was taken as an input. On the other hand, Athanassopoulos (1998) considered competition an attribute, that is, a factor that does not have a "predetermined positive or negative impact on the output produced" (Athanassopoulos, 1998, p. 178). Drake and Howcroft (1995) and Avkiran (1999a) also used competition on the input side of their efficiency assessments, but apparently did not consider this input as negative.

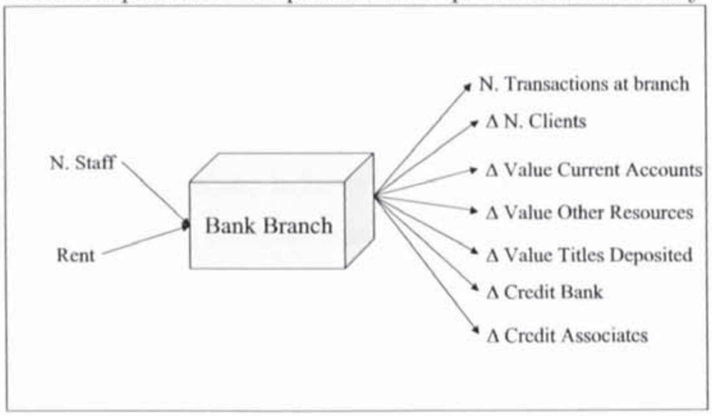
The operational input considered in Figure 5.4 is the number of staff, which is the most important input in the operational activities of bank branches. Other type of inputs such as technology related variables could also have been considered. We did not consider these in our case because the set of bank branches to be analysed are quite homogeneous and they use the same type of technology. Variables concerning number of computers, although available, do not seem to add much to the above assessment because operational efficiency is not so much related to the number of electronic devices each staff member has, but to how the potential of these devices is actually exploited.

The outputs of the operational efficiency assessment reflect the number of new clients in a given period, and the number of sales of various products during the same period. Alongside with selling, staff also serve clients that go to the branch for transactional purposes. This activity is imbedded in the variable 'number of transactions' at the bank branch. Note that, as mentioned before, this output will be considered non-discretionary.

In the set of outputs above we also consider a service quality index. The consideration of this factor goes in line with the notion that branch staff should sell as much as possible but not at the expense of reducing service quality levels. We have however some doubts in using this output (and in fact we did not use it in our actual set of inputs and outputs) because its consideration implies that, for example, increasing staff is likely to increase service quality (positive relationship between inputs and outputs inherent in any efficiency assessment), and that increasing service quality may be done at the expense of reducing some of the others outputs (substitutability between outputs inherent in any efficiency assessment). Obviously we accept that more resources are consumed to produce higher service quality. The question is whether in this assertion we are referring to inputs of the type we can see in Figure 5.4. Is it the number of staff that potentially increases the quality of the service provided or is it their quality? What about the other variables, mostly environmental? Would they have any impact on service quality? On the other hand, on the output side can we say that the bank branch is willing to reduce some sales of accounts to provide better service quality? In theory we could, but in practice things might not work exactly in that way. For this reason, in our empirical application considerations about service quality and its relationships with operational efficiency will be performed *a posteriori*.

Given data availability, the set of inputs and outputs that we use in the operational efficiency assessment is shown in Figure 5.5. The actual set of inputs used in this assessment includes just the number of staff and rent, where the latter is a surrogate for the environment of the branch. The variable relating to the market potential of the branches was not available and therefore it could not be used in the assessment. As far as the other ideal inputs are concerned, namely competition and age, these are available but we decided not to use them. The variable relating to the level of competition that we have is the number of competing branches within one kilometer of each branch. We did not use this variable

Figure 5.5: Actual Inputs and Outputs in the Operational Efficiency Assessment



because the way it was constructed did not seem the most appropriate for inclusion in our study. Indeed, a competition level of 5 branches within one kilometer might mean completely different things depending on the branch being located on an urban area or a rural area (small competition in the first case and high competition in the second case). At the same time doubts arose as wether to consider competition an isotonic or non-isotonic input as from our data it was not possible to infer the type of effect competition had on bank branches' outputs.

Concerning the variable age, its non-inclusion in the assessment relates primarily to the fact that only 4 branches in our sample are younger than four years old. This means that any impact of the newness of the branch had worn off because most of the branches were sufficiently old for that. In addition, if age was to be considered it would be a non-isotonic input (the higher the age the lower the expected amount of outputs), which would create additional difficulties to our analysis.

It is therefore, recognised that age and competition might have an impact on the network of bank branches under analysis, but this effect will only be analysed *a posteriori*.

With reference to outputs we did not have data concerning sales of the various products of the bank branch. However, as we have monthly data we can use the variation in each account from month to month as a *proxy* for sales. Following this approach, we considered on the output side the change in current accounts, in other resources (that include term deposit accounts, emigrant accounts, investment funds, savings insurance, etc.), in titles deposited, in credit by the bank and in credit by associates. We consider here two broad classes of credit (as classified by the bank): credit by the bank includes consumer's credit, card's credit, housing credit, commercial discounts, loans, overdrafts, share's credit, amongst others; and credit by associates includes factoring and leasing. The latter type of credit is called credit by associates because it is not the bank that directly provides it but other companies (associates) inside the same financial group.

The output number of transactions in Figure 5.5 is a sum of various transactions for which we have detailed information. In this sum the bank does not weight differently the various transactions considering therefore that these consume the same amount of resources of bank branches. Note that the transactions considered here relate to all transactions taking place at the bank branch irrespective of the client being from that branch or from another branch of the group. This does not happen with the other outputs that relate only to branch clients. Ideally, the variables capturing the variation in the various accounts should relate to both branch's clients and non-clients, because the operational work load of the bank branch relates to the sales it does irrespective of it being for a branch client or not. The consideration of only branch clients in the computation of change of accounts represents, therefore, a limitation of our empirical analysis that may underestimate the efficiency of some bank branches performing a large number of sales to non-branch clients.

In the ideal inputs and outputs specified in Figure 5.4 output variables are specified in number. However, in the actual operational efficiency measurement we could not use quantity variables because these were not available. We use therefore value information that concerns the amount of money that is kept in each of the various accounts. This fact constitutes an important limitation of our empirical analysis.

The use of changes in activity levels as outputs results necessarily in some outputs being negative for some of the branches. At least two other bank branch studies in the literature have dealt with the same issue. Pastor (1994) considered two outputs in their study of the operational efficiency of a set of 23 Spanish bank branches: Change in demand deposits, and change in time deposits. As both these outputs are unrestricted in sign the authors used a BCC input oriented model that is translation invariant on outputs. Sevcovic et al. (2002) analysed the efficiency of 37 branches and 591 sub-branches offices in Slovakia. One of the outputs considered could be negative and as such the authors adopted the additive translation invariant model of Lovell and Pastor (1995), where slacks are normalised by the standard deviation. The existing approaches for dealing with negative data have some flaws and therefore we developed our own models to handle negative outputs. The developed measures are detailed in Chapter 6.

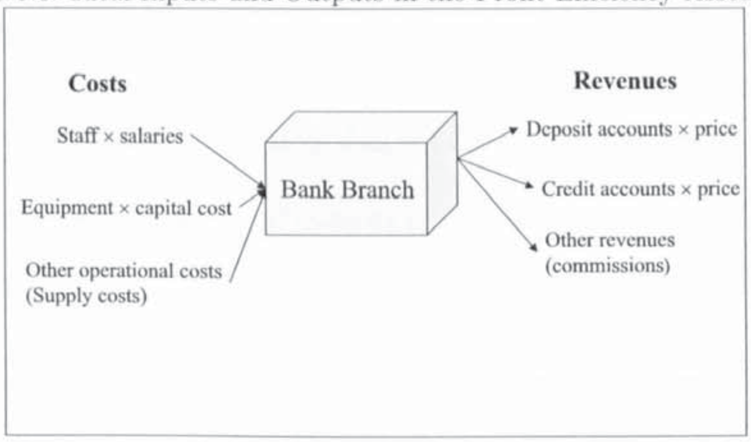
5.2.3 Profit Efficiency

The profit efficiency measure is intended to capture the extent to which a bank branch is profit maximising. Note that our interest here is not simply to calculate what the profit (understood as an accounting concept) was, but to compare the potential for profit maximisation of bank branches.

In the measurement of profit efficiency there are two types of variables that are to be used: values and prices. By values we understand the amount of money that is kept in the various accounts, and by prices we mean interest rates.

The set of ideal inputs and outputs to be used on the profit efficiency assessment can be seen in Figure 5.6. These inputs and outputs are related with the intermediation efficiency

Figure 5.6: Ideal Inputs and Outputs in the Profit Efficiency Assessment



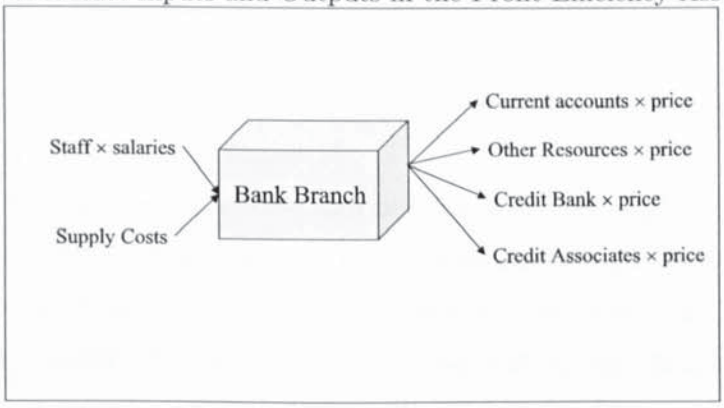
approach in the sense that on the input side costs are considered and on the output side revenues are considered. Note that costs and revenues are disaggregated in value and price information. In this sense, our approach differs from those used in the literature to measure profit (intermediation) efficiency, where prices are not taken explicitly into account. Ignoring prices is equivalent to assuming that these are the same for all production units and that no product is more important than any other in determining revenues. This is clearly a limiting assumption and therefore, whenever possible, price information should be included in a profit analysis. The framework developed to compute profit efficiency is detailed in Chapter 8. The disaggregated information on prices and values will be used in this framework, where the use of value variables alone allows the measurement of technical profit efficiency, and the use of value information together with price information allows the calculation of maximum profit and profit efficiency.

Our specification of outputs in Figure 5.6 considers a set of accounts from which revenue is generated. This means that the sum of our outputs (except commissions) multiplied by their respective prices gives the total interest revenues. In Figure 5.6 we do not consider interest costs on the input side because we are assuming that the prices provided for each output are net interest rates, and therefore the revenue thus obtained is also a net interest revenue (interest revenue - interest costs). This obviously simplifies the model because the products which are a source of interest revenue are also a source of interest cost for the bank (for example, deposits give an interest revenue to the bank from its application on monetary markets but the bank also incurs in an interest cost that is what the client earns).

Note that the specification of inputs and outputs in Figure 5.6 does not require the existence of price information for all inputs and outputs (supply costs and commissions are two cases for which price information might not exist). In Chapter 8 we mention this fact and provide a way to deal with this problem.

The actual set of inputs and outputs that will be used in our empirical application is shown in Figure 5.7. This set of variables does not differ much from the ideal set of inputs

Figure 5.7: Actual Inputs and Outputs in the Profit Efficiency Assessment



and outputs seen in Figure 5.6. The main differences relate to the impossibility of using capital costs on the input side of our efficiency assessment, and to the impossibility of using the value of non-interest revenues (commissions) on the output side. None of this data was available for the bank branches being analysed.

We have price information for all outputs and also for staff. For the input *supply costs* price information is not disaggregated and therefore supply costs shall be included as such in the analysis (in a way that is explained in Chapter 8). Concerning output prices the bank could not supply interest rates per branch, but only average interest rates across all branches. This means that net interest rates of outputs are equal for all bank branches.

Although this is not strictly true it is an approximation to reality. In fact the bank sets prices centrally, leaving bank branches a limited freedom in setting different prices for different customers. Prices can, therefore, vary by client but one may assume that on average they do not differ much between branches. Nevertheless, the ideal situation for measuring profit efficiency would be one where individual price information per branch was available, since the use of the same prices means that the analysis considering output prices is equivalent to one where revenues were considered instead. It was decided, even with the above limitation, to use price and value information disaggregated so that all the models developed are fit to be used in the ideal case where prices are specific of each branch.

5.3 Summary

In this Chapter we put forward a model to assess the efficiency of a set of Portuguese bank branches. Three measures of efficiency are defined in this Chapter, but its application to a sample of bank branches is postponed to Chapter 10. This is because before applying our models, some methodological developments were required so that the above efficiency measures could be in fact used. The methodological developments relate to new efficiency measures that can deal with negative outputs (presented in Chapter 6), a novel approach for providing closest targets to production units (developed in Chapter 7), and a new measure of profit efficiency (developed in Chapter 8).

Finally note that the above proposed measures bring in new insights to the traditional efficiency measurement of bank branches. On the one hand the concept of transactional efficiency is introduced here for the first time, and, on the other hand, the concept of production or operational efficiency is adapted to account for the changing role of bank branches (namely it focus on selling and moves the servicing of clients on general transactions to a second place). In addition, we also consider the main objective of bank branches as for-profit organisations and propose ways to measure their profit efficiency in a way that accounts for both values and prices.

The three objectives of bank branches, and the corresponding efficiency measures, are obviously linked as will be shown in Chapter 11, where we compare results obtained from the three separate efficiency assessments.

Chapter 6

DEA Models to Measure Operational Efficiency in the Presence of Negative Data

In this Chapter we present a novel approach to deal with negative data. Such approach was inspired by the directional distance function and has a number of advantages over existing approaches to deal with this type of data.

6.1 Introduction

Data Envelopment Analysis (DEA) is an efficiency assessment tool that implicitly assumes positivity of all inputs and outputs. However, in many real life contexts not all inputs and outputs are positive for all operating units. As noted in the previous Chapter, in our case too the measurement of operational efficiency, where outputs refer to changes in the levels of clients, deposits, etc., resulted in some outputs taking negative values.

Traditionally negative data are handled in efficiency applications through some data transformation (e.g. adding an arbitrary large number to all values of a given variable) so that all negative data are turned into positive data (see e.g. Pastor, 1994; Lovell, 1995). Such transformation of the data may have implications for the solution, classification, or ordering of the DEA results (Seiford and Zhu, 2002). There are, however, some models whose solution is invariant to data transformations, which are usually referred to as translation invariant. In the presence of negative data the most often used model is the variable returns to scale (VRS) additive model of Charnes et al. (1985b), which is translation invariant as demonstrated by Ali and Seiford (1990). The additive model is not however, in its original form, units invariant (independent of scale of measurement of the variables). Due to this limitation Lovell and Pastor (1995) put forward a units invariant version of the additive model that uses a weighted sum of slacks where the weights are the inverse of the standard deviations of the corresponding input and output variables (see also Pastor, 1996; Thrall, 1996). The main advantage of the additive model is that it can be applied to negative data directly without any need to subjectively transform them. However, the additive model has some drawbacks, namely the fact that it yields in respect of an inefficient unit the 'furthest' targets on the production frontier, while at the same time it does not yield an efficiency measure that can be readily interpreted. Thus the model does not yield very practical guidance as to how a unit might improve its performance nor does it make it possible to readily rank units on performance. The VRS model of Banker et al. (1984) (also known as BCC model) is able to provide an efficiency score in the presence of negative data, but this cannot be achieved without transforming the data. In addition, the BCC model has restricted translation invariance (it is translation invariant on inputs if it is output oriented, and translation invariant on outputs if it is input oriented (see e.g. Lovell and Pastor, 1995; Pastor, 1996).) meaning that the efficiency scores may depend on the way data are translated.

Thus there is no DEA model to date that can be used with negative data directly without any need to transform them while at the same time it yields an efficiency score that can be readily used to compare units. In this Chapter we propose DEA models which provide efficiency scores, similar in meaning to radial efficiencies traditionally used in DEA, while at the same time negative data can be used without the need to subjectively transform them. This is an important advantage over existing approaches to deal with negative data.

Our approach is inspired by the well known directional distance model of Chambers et al. (1996a, 1998), and it provides efficiency scores that can be directly used to rank and compare production units when some inputs and/or outputs are negative. Targets resulting from our procedure, and from a variant of this procedure, are also analysed in this Chapter and it is shown that our models in general provide closer targets than existing models in the literature. Closer targets represent a useful practical feature because they would prove easier for the unit to attain and have been explored for the case when all data are positive in a number of papers (e.g. Coelli, 1998; Frei and Harker, 1999; Cherchye and Puyenbroeck, 2001a; Portela et al., 2003, amongst others).

We assume throughout that negative data is not necessarily a bad in itself, i.e. efficiency does not require all inputs and outputs to be positive. Rather we assume that it is possible to have certain efficient units presenting some negative values of inputs and/or outputs. Consider for example the case of our bank branches where negative outputs arise from the fact that we are measuring growth on the output side. If we have, for example, one output measuring sales growth and another measuring clients growth, it is likely that a bank branch consciously chooses to increase more one of the outputs at the expense of reducing the other. That is in managing its operations a branch may prefer to increase sales by selling more to existing customers rather than to increase the customer base so that the basis for potential sales increases. These two strategies are equally valid although they may mean in the short run neglecting one of the factors in favour of the other. The neglecting factor may thus suffer a decrease, and growth will be negative.

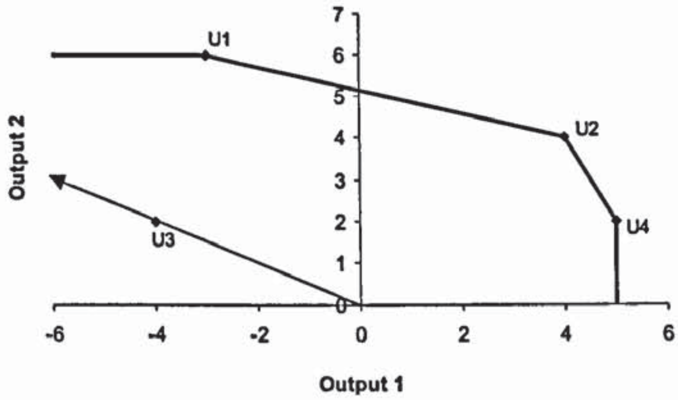
6.2 Negative Data: Implications in DEA

Negative data may arise due to the consideration of variation in variables like changes in clients or accounts from one period to the other (Pastor, 1994), or due to the use of variables like profit that may take both positive and negative values (Krivonozhko et al., 2001). Negative inputs or outputs may also arise artificially as a way to deal with undesirable inputs or outputs (Seiford and Zhu, 2002).

The translation invariance of the additive model is subject to it being specified under VRS. Constant returns to scale (CRS) models are not translation invariant. In fact the notion of CRS is undefined in the presence of negative data. We can readily demonstrate that an assumption of CRS, as traditionally defined, is not possible in technologies where negative data can exist. A CRS technology assumes that any activity can be "radially expanded or contracted to form other feasible activities" (Färe et al., 1994a, pg. 50). Take a set of only two units, A and B, represented by activity vectors (x, y_1, y_2), where x is input and y_1 and y_2 outputs. Assuming that output 1 is negative, consider that A and B equal (1, -1, 1) and (1, -2, 3), respectively. Unit A has higher productivity in y_1 and B has higher productivity in y_2, and therefore both units are CRS efficient (see Chen and Ali, 2002). However, the additive CRS model shows only unit B efficient. In fact, it is possible to radially contract unit B (say by 50%) and find a feasible point (under the CRS assumption) dominating unit A [e.g. 0.5B =(0.5, -1, 1.5) dominates A]. The productivity ratios, however, remain unchanged and as such this result is clearly wrong. The assumption behind CRS DEA models that any proportion of an efficient unit is also efficient, is therefore only valid for non-negative data.

In the presence of negative data VRS technologies need to be assumed. However, the use of radial measures of efficiency traditionally used in VRS DEA models is problematic. To illustrate the point consider the example in Figure 6.1, where two outputs are represented (output 2 is positive and output 1 may be negative) and all units have the same input. Assessing the efficiency of unit U3 using, for example, the radial output oriented BCC

Figure 6.1: Example with one negative output



model (Banker et al., 1984) without transforming the data, implies an expansion of both outputs by a multiple greater than 1. This, however, implies a movement of the inefficient unit U3 to the frontier in the direction shown by the arrow in Figure 6.1. This movement is not desired since the negative output is being expanded making it even worse. Clearly positive radial expansion factors applied to negative data lead in the opposite direction to the one we would wish to follow to improve performance. The addition of a constant to the negative output (output 1 in Figure 6.1) would move the frontier to the positive quadrant and the right direction would be followed by U3 towards the frontier. The output efficiency score would, however, depend on the value of the constant added to the negative output vector, and the resulting radial efficiency score would be hard to interpret in the light of the negative data it in fact represents.

Note that the treatment of negative data is in a way similar to the treatment of undesirable inputs and/or outputs, since both negative data and undesirable outputs need to be constrained to move in a direction that is contrary to the direction used in traditional DEA models. Several approaches exist to deal with undesirable outputs as can be seen in the recent review of Allen (1999) and Dyckhoff and Allen (2001). One of these approaches is based on the directional distance function, and was first proposed by Chung et al. (1997). In this Chapter we use a related approach, also based on the directional distance model, to deal with negative data.

6.3 A Directional Approach to Deal with Negative Data

Consider a set of units $j = 1, \ldots, n$, with input levels x_{ij}, $i = 1, \ldots, m$ and output levels y_{rj}, $r = 1, \ldots, s$, and unit $o \in j$ which is to be assessed. The generic directional distance model as proposed by Chambers et al. (1996a, 1998) is in (6.1) for the case of VRS and with input and output vectors in R^{m+s}.

$$\max\left\{\beta_o \mid \sum_{j=1}^{n} \lambda_j\, y_{rj} \geq y_{ro} + \beta_o\, g_{y_r}\ r = 1, \ldots, s,\ \sum_{j=1}^{n} \lambda_j\, x_{ij} \leq x_{io} - \beta_o\, g_{x_i}\ \ i = 1, \ldots, m, \right.$$
$$\left. \sum_{j=1}^{n} \lambda_j = 1,\ \ \lambda_j,\ \beta_o,\ g_{x_i},\ g_{y_r}\ \geq 0\right\} \quad (6.1)$$

Model (6.1) is defined in the most general non-oriented case as it looks simultaneously for input contraction and output expansion. Oriented models can be derived from (6.1) by setting respectively g_{y_r} or g_{x_i} equal to zero. When data are strictly positive a usual choice for the directional vectors (g_{x_i}, g_{y_r}) are the observed input and output levels. When some data are negative, the use of observed input and output levels would violate the last constraint of model (6.1), which is intended to ensure that inputs and outputs do not worsen from their observed levels in the solution the model yields.

We modify the model in (6.1) to ensure it yields improving solutions even when some of the data are negative. Specifically, and for a given data set, consider an ideal point defined as $I = (\max_j \mathrm{y_j},\ r = 1, \ldots, s,\ \min_j \mathrm{x_j},\ i = 1, \ldots, m)$. We can now define the vectors R_{ro} and R_{io} in (6.2), to which we refer as the **range of possible improvement of unit** o.

$$R_{ro} = \max_j \{y_{rj}\} - y_{ro},\ r = 1, \ldots, s \ \text{ and }\ R_{io} = x_{io} - \min_j \{x_{ij}\},\ i = 1,\ \ldots, m \qquad (6.2)$$

Although there is no evidence that any unit can actually exist at the *ideal* point I the range of possible improvement in (6.2) can be seen as a surrogate for the maximum improvement that unit o could in principle achieve on each input and output. Such an improvement can never be negative, and therefore the range vectors in (6.2) satisfy the non-negativity restrictions on the direction vectors used in (6.1). Under VRS units that have the maximum value on some output or the minimum value on some input are always 100% efficient (Chen and Ali, 2002). Thus the range of possible improvement we use is determined by the efficient units' input/output levels, which is already a characteristic inherent in the classical DEA model (e.g. Thanassoulis, 2001, Chapter 3).

Note that this contrasts with other notions of 'range' used in the literature such as by Cooper et al. (1999), where range of a variable is defined as its maximum observed minus its minimum observed value. In such a range worst performance as given by maximum inputs and minimum outputs affects the results of the model. This is because worst performance is included in the definition of the range and efficiency results depend on the range defined. Another notion of 'range' related with that defined in (6.2) has been introduced by Bogetoft and Hougaard (1998) and also used by Asmild et al. (2003). Bogetoft and Hougaard (1998) introduce a 'potential improvements approach' using the input oriented directional distance function, where the directional input vector is the difference between the observed input and an ideal reference input. This ideal input vector, however, is specific to each production unit reflecting the "largest possible reduction in each input with all other inputs kept fixed" (Bogetoft and Hougaard, 1998, p. 235). To the authors' knowledge the use of the range direction as specified in (6.2) has never been used before in the literature.

6.3.1 Range Directional Model (RDM)

Based on the notion of the range of possible improvement in (6.2), we define the Range Directional Model (RDM) as shown in (6.3).

$$\max\left\{\beta_o \mid \sum_{j=1}^{n}\lambda_j\, y_{rj} \geq\, y_{ro}+\beta_o\, R_{ro}\;\; r=1,\ldots,s,\;\; \sum_{j=1}^{n}\lambda_j\, x_{ij} \leq\, x_{io}-\beta_o\, R_{io}\;\; i=1,\ldots,m \right.$$
$$\left. \sum_{j=1}^{n}\lambda_j=1,\;\lambda_j\geq 0\right\} \quad (6.3)$$

Models that can deal with negative data should be translation invariant and units invariant. The RDM in (6.3) satisfies both these properties as proved next.

Translation Invariance Proof: Let an amount K_r be added to each output and V_i to each input. The constraints in (6.3), therefore, become: $\sum_{j=1}^{n}\lambda_j\;(y_{rj}+K_r)\geq(y_{ro}+K_r)+\beta_o\;R_{ro}$ and $\sum_{j=1}^{n}\lambda_j\;(x_{ij}+V_i)\leq\;(x_{io}+V_i)-\beta_o\;R_{io}$. Note that the range of improvement does not change with the addition of a constant to each input and output. The left hand side of the output inequality ($\sum_{j=1}^{n} \lambda_j \ (y_{rj} + K_r)$) is equivalent to $\sum_{j=1}^{n} \lambda_j \ y_{rj} + K_r \sum_{j=1}^{n} \lambda_j$. As the $\sum_{j=1}^{n} \lambda_j = 1$, then the constraints changed with K_r reduce to the constraints in model (6.3). The same happens with the input constraints changed by V_i.

Units Invariance Proof: Consider that all levels of input i are multiplied by α_i, and of output r by γ_r. This results in the following modified constraints of (6.3): $\sum_{j=1}^{n} \lambda_j(\gamma_r y_{rj}) \geq \gamma_r y_{ro} + \beta_o(\gamma_r R_{ro})$ and $\sum_{j=1}^{n} \lambda_j(\alpha_i x_{ij}) \leq \alpha_i x_{io} - \beta_o(\alpha_i R_{io})$. These constraints reduce to those in (6.3), whose solution, therefore, does not change when the unit of measurement changes.

The translation invariance of the RDM model means that it can be equivalently applied on the original data, where some inputs and/or some outputs may be negative, or on transformed data. Obviously the use of original data is preferred, as data transformations result in the addition of un-necessary complexity to our models. Note that the RDM model is only translation invariant when VRS are assumed, as all the existing translation invariant models in the literature.

The range of improvement R_{ro} or R_{io} as defined in (6.2) may be zero for some outputs and some inputs. This is in line with intuition, because a range of zero improvement means that the unit has achieved on that variable a good enough value so that we have no observed evidence how that value might improve even further. Note that a constraint in (6.3) associated with a zero range is necessarily binding (target values equal observed values).

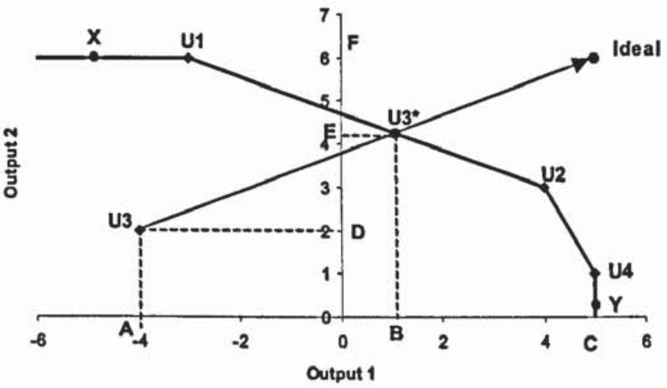
Closer Look at the Efficiency Measure Yielded by the RDM Model

At the optimal solution to model RDM at least one constraint is binding, meaning that β equals $\frac{y_r^* - y_{ro}}{R_{ro}}$ or $\frac{x_{io} - x_i^*}{R_{io}}$ for at least an output r or an input i. The star stands for the target value obtained at the optimal solution to model (6.3). This means β is equal to the ratio of an optimal slack (that projects unit o on the frontier) to the maximum possible slack (given by the range) unit o had on that variable. Seen in this way β is clearly an **inefficiency measure**. The RDM **efficiency measure**, $1 - \beta$, is therefore defined as $\frac{Max_j\{y_{rj}\} - y_r^*}{Max_j\{y_{rj}\} - y_{ro}}$ if a binding constraint corresponds to output r, or $\frac{x_i^* - Min_j\{x_{ij}\}}{x_{io} - Min_j\{x_{ij}\}}$ if a binding constraint corresponds to input i. As target outputs (target inputs) cannot be lower (higher) than observed outputs (observed inputs), the numerator of $1 - \beta$ is never larger than the denominator, meaning that the upper bound of $1 - \beta$ is 1. Efficiency of 1 will only be achieved when the observed are also the target values, i.e. when the unit

o, being assessed, lies on the efficient frontier. This is as in the case of traditional DEA models (Charnes et al., 1978).

The RDM efficiency measure can be illustrated with the aid of Figure 6.2 (depicting the same units as Figure 6.1), where we are assuming an output oriented RDM model. The efficiency measure $1 - \beta$ of U3 equals the ratio $\overline{CB}/\overline{CA}$, which in turn equals the

Figure 6.2: RDM in a 2 output example

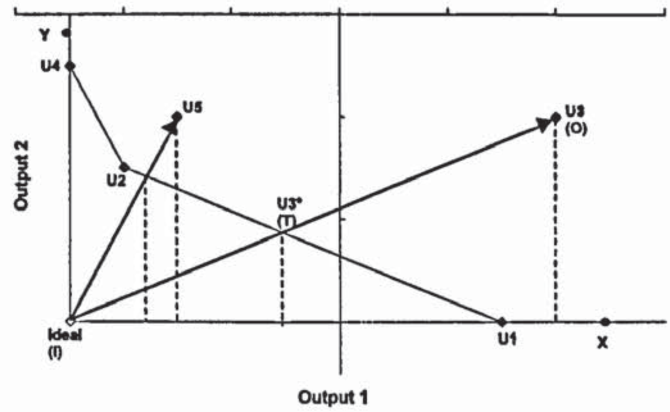


ratio $\overline{FE}/\overline{FD}$. Note that $\overline{CB}/\overline{CA}$ measures the distance between the level of output 1 at the observed point U3 and its target point U3*. $\overline{FE}/\overline{FD}$ is interpreted in a similar manner in respect to the level of output 2. Thus we have for U3 a value of $1 - \beta$ equal to $\frac{5-1.07273}{5-(-4)} = \frac{6-4.25455}{6-2} = 43.36\%$, reflecting the relative distance between U3 and its target $U3^*$.

Note that there is close similarity between the RDM efficiency measure and radial measures of efficiency traditionally used in DEA. The difference is in the reference point used to measure efficiency. In the RDM case the reference point is not the origin used in traditional DEA models but rather the *ideal* point we defined using (6.2). In fact if we rotate Figure 6.2 suitably we can arrive at Figure 6.3 in which the ideal point occupies the position of the origin in traditional DEA models. Using Figure 6.3 it is easy to see that the efficiency measure yielded by model RDM, $1 - \beta$, is a distance measure between the observed and its target point with reference to the ideal point. The lower this distance the higher the value of $1 - \beta$ and the more efficient a unit will be. To see this note that the direction of improvement followed by inefficient units U3 and U5 in Figure 6.3 is defined with reference to the ideal point, a role played by the origin in traditional DEA models.

Our efficiency measure has the same geometric interpretation as radial measures in DEA provided the ideal point is treated as the origin. Consider for example U3 and define

Figure 6.3: Figure 6.2 after rotation



two vectors $\vec{A} = \overrightarrow{I - T}$ that goes from ideal (I) to target (T) point, and $\vec{B} = \overrightarrow{I - O}$ that goes from ideal to observed (O) point. Then the efficiency measure $(1 - \beta)$ of U3 is given by the ratio between the length of these two vectors, that is by $||A||/||B|| = ||I - T||/||I - O||$, exactly as would be under traditional DEA, had the point I been the origin. (For example, assuming an input oriented DEA model, where data are assumed to be non-negative, consider the ratio $||T||/||O||$ where (T) is an input target vector and (O) is an input observed vector. Given T is the radial input target vector it equals θO, where θ is the input efficiency score, meaning that the above ratio of norms reduces to θ.)

Pareto-Efficiency

Efficient units will necessarily have in the RDM model an optimal $1 - \beta$ equal to 1, but this is not a sufficient condition for Pareto-efficiency. The Pareto-optimality conditions for model RDM are two: (i) $\beta_o = 0$, and (ii) all constraints of (6.3) are satisfied in equality (i.e. all slacks are zero). Note that though the RDM model does not assure projection on Pareto-efficient targets it may in some cases correctly identify weak efficiency. This is the case for units X and Y in Figure 6.2 (or in Figure 6.3), whose assessment through the RDM model yields a value of $1 - \beta$ different from 1. This is an interesting characteristic of the RDM model that, though behaving as a radial model, can in some cases identify weak efficient units. Units located on an inefficient part of the frontier have at least one of the ranges equal to zero, which has no influence over the value of β. As β is maximised in model RDM (see (6.3)) any inefficiencies in the factors that have non-zero range may be found because these will push β to be greater than zero. Nevertheless, with several inputs/outputs $\beta = 0$ is not a sufficient condition to assure Pareto-efficiency[1]

Note that to find Pareto-efficient targets one can solve (6.3) in a first stage and the additive model in a second stage as described in Ali and Seiford (1993b). Alternatively Pareto-efficient solutions can be found by solving the additive model in a first stage to identify those units that are Pareto-efficient and then restrict the reference set in the RDM model to those Pareto-efficient units. In our empirical implementation detailed in Chapter 10 we have chosen the latter approach.

The aforementioned implies that the efficiency score $1-\beta$ is not able to incorporate all the sources of inefficiency, since slack values are not reflected in the value of β. Ranking units based on an efficiency measure that does not include all the sources of inefficiency may result in a biased ranking especially if slacks are high. We can, however, use the ratio of the norms ($||A||/||B|| = ||I-T||/||I-O||$, as defined previously) to account for all inefficiencies including those from slacks as long as target levels considered in the computation of A and B are Pareto-efficient (see e.g. Cherchye and Van Puyenbroeck, 1999a,b). Note however, that when Pareto-efficient targets are used in A and B these two vectors are not necessarily collinear, meaning that the resulting efficiency measure is dependent on units of measurement. In order to avoid this problem the ratio of norms ($||A||/||B||$) should be used on normalised data only.

6.3.2 Target Setting under Negative Data in DEA

In the RDM model the direction towards the production frontier is in a sense 'biased' towards the factors with the largest potential for improvement. That is, the model seeks targets such that the factors on which the unit has the largest difference from the 'best' values observed elsewhere are those where improvement is given priority. Thus in a sense the model seeks targets so that the unit will improve in those factors where it does 'worse' relative to other units, and therefore the targets may prove hard for the unit to achieve in the short-run.

This section puts forth an alternative direction of improvement of inputs and outputs so that the unit will identify targets where the factors on which it does best are given priority to improve. Such targets will normally prove easier for the unit to attain in the short term.

[1] Consider, for example, two points W and Z with input/output vectors (y, x_1, x_2) equal to (12, 5, 13), and (12, 5, 17), respectively. Unit W dominates unit Z, which is not Pareto-efficient. However, Z has an optimal β_Z equal to 0 as resulting from model (6.3), where the range for improvement vector is (19, 0, 11). This model has therefore multiple optimal solutions where β_Z is zero both when $\lambda_Z = 1$ and $\lambda_W = 1$.

This direction uses the inverse of the ranges in (6.2) in the context of model (6.3). The resulting model, referred to as Inverse RDM (IRDM), is in (6.4).

$$\max\left\{\beta_o \mid \sum_{j=1}^{n}\lambda_j\, y_{rj} \geq\ y_{ro}+\beta_o\,\frac{1}{R_{ro}}\ r=1,\ldots,s,\ \sum_{j=1}^{n}\lambda_j\, x_{ij} \leq\ x_{io}-\beta_o\,\frac{1}{R_{io}}\ i=1,\ldots,m\right.$$
$$\left.\sum_{j=1}^{n}\lambda_j=1,\ \lambda_j\geq 0\right\} \quad (6.4)$$

For ranges in (6.4) which are zero, division by zero is avoided and we use zero as the coefficient of the corresponding $1/R_{io}$ or $1/R_{ro}$. This treatment of zero ranges ensures that the corresponding input or output has within the targets derived the same value as that observed at the unit concerned. This matches the treatment of zero ranges in the RDM model.

Model (6.4) is translation invariant (the proof for translation invariance of (6.4) is exactly the same as for model (6.3), because the range for improvement does not change when the same constant is added to inputs and/or outputs). However, model (6.4) is not units invariant. Assuming for example that all levels of output r are multiplied by γ_r, we have $\sum_{j=1}^{n}\lambda_j\ \gamma_r y_{rj} \geq\ \ \gamma_r y_{ro}+\beta_o\ \frac{1}{\gamma R_{ro}}$, which is not equivalent to the constraint for output r in (6.4). In order to circumvent this problem we can use normalised data in model (6.4), so that its solution is not dependent on unit of measurement. Using normalised data, and ranges calculated on this normalised data, makes model (6.4) units invariant (As $\sum_{j=1}^{n}\lambda_j\frac{y_{rj}}{Y_r} \geq \frac{y_{ro}}{Y_r}+\beta\frac{Y_r}{R_{ro}}$ (where Y_r is the maximum output r), is equal to $\sum_{j=1}^{n}\lambda_j\frac{\alpha y_{rj}}{\alpha Y_r} \geq \frac{\alpha y_{ro}}{\alpha Y_r}+\beta\frac{\alpha Y_r}{\alpha R_{ro}}$. Note that the same is valid for input constraints.) **The IRDM model is, therefore, translation and units invariant on normalised data.** In our empirical application, detailed in Chapter 10 model IRDM is applied to data that have been previously normalised by a non-negative value.

Model (6.4) must be used in our empirical application for target setting purposes only. This is because the "efficiency measure" it yields does not have a straightforward interpretation as will be seen next.

The IRDM efficiency score $1-\beta$ in (6.4) measures the distance from an observed point to a target point with reference to some ideal point. However, the IRDM model works as if a different ideal point was defined for each unit, which represents a problem in interpreting and comparing efficiency scores yielded by this model[2].

[2]To illustrate this consider the optimum β, as resulting from (6.4), after normalising outputs by Y_r and

Since the IRDM model works as if different ideal points were defined for each production unit under assessment, the IRDM efficiency measures are not comparable within themselves nor with RDM efficiency scores. This means that the IRDM model should not be used to rank and compare units but just for target setting purposes.

6.3.3 Closest Targets and the RDM Models

The IRDM model gives priority to improve the factors on which production units perform best. As a result one expects targets derived from this model to be less demanding (closer) than those resulting from the RDM model. The IRDM model may be, therefore, a good alternative to more complicated procedures of finding closest targets to inefficient units. For example, the procedure developed by Charnes et al. (1992) for calculating the radius of stability (minimum change needed to change the classification of a unit) can be used for calculating targets with the minimum L_1 distance. This procedure is not, however, units invariant and implies solving several linear programming models $(m + s)$, each being an additive model that maximises slack variables in turn (see also Briec (1998) who put forward the same model for finding the minimum L_1 projection). In Chapter 7 we show a procedure for finding closest targets to inefficient units. This procedure is, however, based on a measure that cannot be directly applied in the presence of negative data as it is based on ratios of target to observed input or output levels (which would be meaningless in case observed data are negative).

The IRDM model does not assure closeness on any criteria (such as any L_p metric) but by focusing improvements on the factors at which the unit is already good at it provides in principle targets that are near the closest.

We illustrate this point through the example that has been used previously (see e.g. Figure 6.2), namely showing distances from unit U3 to alternative targets. We consider 6 different targets to unit U3 in Table 6.1: target U1 (-3, 6) is the closest target to this unit according to the procedure of Charnes et al. (1992); target U2 (4, 3) results from

inputs by X_i, which is $\frac{(y_r^* - y_{ro})/Y_r}{Y_r/R_{ro}}$ or $\frac{(x_{io} - x_i^*)/X_i}{X_i/R_{io}}$ when respectively the constraint relating to output r or input i is binding at the optimal solution to the IRDM model. Consider now a new range of improvement as given by $R'_{ro} = \frac{Y_r}{R_{ro}}$ for outputs, and $R'_{io} = \frac{X_i}{R_{io}}$ for inputs, and normalised target and observed levels equal to $y_r'^* = y_r^*/Y_r$, $y'_{ro} = y_{ro}/Y_r$, $x_i'^* = x_i^*/X_i$, and $x'_{io} = x_{io}/X_i$. The optimal value of β in the IRDM model reduces therefore to $\frac{y_r'^* - y'_{ro}}{R'_{ro}}$ or $\frac{x'_{io} - x_i'^*}{R'_{io}}$ when respectively the constraint relating to output r or input i is binding at the optimal solution to the IRDM model. The above IRDM efficiency relates directly with the RDM efficiency measure, where the range of possible improvement is defined in relation to an ideal point $I' = (y'_{ro} + R'_{ro}, x'_{io} - R'_{io})$. Point I' is no longer fixed as it was in the case of the RDM model, but varies for each production unit.

Table 6.1: Distance of U3 from Some Targets

Target	L_1	L_2	L_∞	$\|\|A\|\|/\|\|B\|\|$
Unit U1	5	$\sqrt{17}$	4	83.35%
Unit U2	9	$\sqrt{65}$	8	28.05%
BCC1	5.47	$\sqrt{16.63}$	3.65	74.84%
BCC2	7.75	$\sqrt{37.54}$	5.81	37.73%
RDM Tgt	7.33	$\sqrt{30.82}$	5.07	43.64%
IRDM Tgt	5.12	$\sqrt{16.75}$	3.9	81.20%

solving the translation invariant additive model of Lovell and Pastor (1995); target BCC1 (-2.1765, 5.65) results from solving the BCC model, where output 1 is transformed into a positive output by adding 5; target BCC2 (1.8125, 3.9375)[3] results from solving the BCC model, where output 1 is transformed into a positive output by adding 10; target RDM (1.073, 4.255) results from the RDM model; and target IRDM (-2.793, 5.911) results from the IRDM model.

U1 is the target yielding the smallest L_1 norm, but this is not true for the other L_p metrics, where the IRDM target performs better than the procedure of Charnes et al. (1992). Note that the BCC1 target performs very well in most L_p norms except in the L_1. Note also that the translation of the data has a big impact on the target levels obtained and also on their distance from observed levels.

Results in terms of L_p metrics should, however, be interpreted carefully because these metrics are units dependent. This means that they are only valid when variables are measured on the same scale or else they can induce completely wrong interpretations. The units independent ratio of norms ($||A||/||B|| = ||I-T||/||I-O||$) can be used to calculate the distance between any observed vector and a target vector with reference to the ideal point[4]. For the alternative targets shown in Table 6.1 we calculated the ratio of norms based on values normalised by the maximum and used a common ideal point as defined by

[3]Note that the target levels for output 1 obtained directly from the BCC model are always non-negative due to the transformation imposed, but we then re-transformed output 1 targets by subtracting a value of 5 and 10, respectively for BCC1 and BCC2.

[4]The ratio of norms is not units invariant unless the two vectors A and B are collinear. The vectors A and B are collinear in the RDM case when the targets are given by $y_{ro}+\beta^* R_{ro}$ or $x_{io}-\beta^* R_{io}$. We call these radial targets as they only expand outputs or contract inputs by the optimal value of β^* as resulting from the RDM model. In the IRDM case collinearity between A and B happens when 'radial' normalised targets are considered. Collinearity between two vectors $A=(a_1,a_2,a_3,\ldots)$ and $B=(b_1,b_2,b_3,\ldots)$ implies that $\frac{a_1}{b_1}=\frac{a_2}{b_2}=\frac{a_3}{b_3}=\ldots$. To prove the units invariance of collinear vectors assume that the units of measurement of a given variable have changed by α. The resulting ratio of norms (in a three dimension case) is equal to

$$\frac{\sqrt{a_1^2+a_2^2+(\alpha a_3)^2}}{\sqrt{b_1^2+b_2^2+(\alpha b_3)^2}}$$

maximum outputs [I = (5, 6)]. This means that all the ratios of norms in Table 6.1 are comparable amongst themselves. The highest value for this ratio happens for projection of U3 on U1, with the second best being attributed to the IRDM target. Note that the additive target (U2) shows the highest distances from targets in all the criteria, and the RDM target lies somewhere between the IRDM and the additive model's targets.

The above results are only illustrative, but they support our argument that the IRDM model has the advantage of looking for closer projections on the efficient frontier when compared to the RDM or to the additive model. The empirical application presented in Chapter 10 will further deal with the issue of closest targets.

6.4 Non-Discretionary Factors in the RDM Model

The RDM model developed so far considered the most general case of non-oriented efficiency measures. In cases where some factors are not under the control of the units being assessed one may need to use oriented models. As mentioned earlier, in order to consider oriented RDM models the only change required to models (6.3) and (6.4) is to set the input directional vector to zero for output oriented models, or set the output directional vector to zero for input oriented models. The changes required for dealing with non-discretionary factors are similar if we adopt the approach of Banker and Morey (1986a). That is, for outputs $r \in NDO$, where NDO is the set of non-discretionary outputs, we set in models (6.3) and (6.4) R_{ro} and $\frac{1}{R_{ro}}$ to zero. For inputs $i \in NDI$, where NDI is the set of non discretionary inputs, we set in models (6.3) and (6.4) R_{io} and $\frac{1}{R_{io}}$ to zero.

Other treatments of non-discretionary factors (as detailed in section 2.3.3 of Chapter 2) can also be adapted to be used in the RDM models (e.g Banker and Morey, 1986b; Ruggiero, 1998; Ray, 1991; Fried et al., 1999). We will adopt the Banker and Morey (1986a) procedure in our empirical application to bank branches, and therefore we will not provide details on other approaches to treat non-discretionary factors under the RDM models.

Being A and B collinear we can replace a_1 by $\frac{a_3 b_1}{b_3}$, and a_2 by $\frac{a_3 b_2}{b_3}$, which results in

$$\frac{\sqrt{(b_1 \frac{a_3}{b_3})^2 + (b_2 \frac{a_3}{b_3})^2 + (\alpha b_3 \frac{a_3}{b_3})^2}}{\sqrt{b_1^2 + b_2^2 + (\alpha b_3)^2}} = \frac{a_3}{b_3} = \frac{a_2}{b_2} = \frac{a_1}{b_1} = \frac{||A||}{||B||}$$

When A and B are not collinear (which happens in the IRDM procedure when we take a fixed ideal point, and in the RDM and IRDM procedures when we take Pareto-efficient targets rather than 'radial' targets), then the ratio of norms calculated on normalised variables shall be used so that the resulting value is not dependent on units of measurement. The ratio of norms can therefore be used to compare alternative target points on the frontier, as long as they are based on normalised values.

6.5 Dual RDM Model

In our application of the RDM model to bank branches (see Chapter 10) we had negative outputs resulting from differences between values observed in two successive time periods. Constructing the dual of the RDM model it can be shown that the use of such negative data is equivalent to a weight restricted model that uses only positive and ratio scale data. In order to illustrate this consider the RDM model as shown in (6.3), whose dual is shown in (6.5).

$$\min\left\{h_o=\sum_{i=1}^{m} v_i\, x_{io}-\sum_{r=1}^{s} u_r\, y_{ro}+u_o \mid \sum_{i=1}^{m} v_i\, x_{ij}-\sum_{r=1}^{s} u_r\, y_{rj}+u_o\geq 0,\right.$$
$$\left. j=1,\ldots,n,\ \ \sum_{r=1}^{s} u_r\, R_{ro}+\sum_{i=1}^{m} v_i\, R_{io}\geq 1,\ \ r=1\ldots,s,\ \ u_o \text{ is free}\right\} \tag{6.5}$$

The output values y_{rj} in model (6.5) are free in sign, and are equal to $y_{rj_t}-y_{rj_{t-1}}$, where t refers to time (in our case months) (see Chapter 5, section 5.2.2 where we present the actual inputs and outputs that are to be used in the operational efficiency assessment). Therefore replacing $y_{rj}=y_{rj_t}-y_{rj_{t-1}}$ in model (6.5) we arrive at model (6.6), where all outputs are positive.

$$\min\{h_o=(\sum_{i=1}^{m} v_i\, x_{io}+\sum_{r=1}^{s} u_r\, y_{ro_{t-1}})-\sum_{r=1}^{s} u_r\, y_{ro_t}+u_o \mid$$
$$(\sum_{i=1}^{m} v_i\, x_{ij}+\sum_{r=1}^{s} u_r\, y_{rj_{t-1}})-\sum_{r=1}^{s} u_r\, y_{rj_t}+u_o\geq 0,\ \ j=1,\ldots,n,$$
$$\sum_{r=1}^{s} u_r\, R_{ro}+\sum_{i=1}^{m} v_i\, R_{io}\geq 1,\ \ r=1\ldots,s,\ u_o \text{ is free}\} \tag{6.6}$$

As can be seen in model (6.6) the variables $y_{rj_{t-1}}$ behave exactly as inputs. We can therefore consider this variable on the input set as long as we add a further constraint imposing that the weights on these inputs equal the weights on the corresponding output. If we consider on the input side, say, clients in February and on the output side clients in March, then we must impose an additional constraint stating that the weight of clients February should be equal to the weight of clients in March. This means that model (6.6) reduces to model (6.7), where the set of inputs m was enlarged to M by adding s inputs

$y_{rj_{t-1}}$.

$$\min\{h_o = \sum_{i=1}^{M} v_i\, x_{io} - \sum_{r=1}^{s} u_r\, y_{ro_t} + u_o \mid$$

$$\sum_{i=1}^{M} v_i\, x_{ij} - \sum_{r=1}^{s} u_r\, y_{rj_t} + u_o \geq 0,\ \ j = 1,\ldots,n,\ \ \sum_{r=1}^{s} u_r\, R_{ro} + \sum_{i=1}^{m} v_i\, R_{io} \geq 1,\ r = 1\ldots,s$$

$$u_r = v_i,\ \text{for}\ \ r = 1,\ldots,s \text{ and } i = m+1,\ldots,M,\ u_o \text{ is free}\} \quad (6.7)$$

Model (6.7) is very similar to other DEA models and assumes positivity of all variables. It imposes, however, some weight restrictions namely those relating with the equality of some input and output weights, and the constraint imposing the weighted sum of ranges of possible improvement to be greater than 1. Note that in this last constraint the ranges considered are still the same as in the original model, where improvement is calculated based on the original y_{rj} and not on each of its components $y_{rj_{t-1}}$ and y_{rj_t}. This happens because, although we can express y_{rj} as the difference between values observed in two successive time periods, we cannot express the range of possible improvement R_{ro} as the difference between ranges of possible improvement in two successive periods.

These results show that the use of variation or interval data (unrestricted in sign) is in fact equivalent to using the original positive values at the beginning and end of a month. This reasoning goes in line with the discussion raised recently in Halme et al. (2002). These authors argue that "in most cases the negative observations in the data result from the fact that the input and output is measured on an interval scale. Since radial DEA models require ratio scale data, the problem is not the negative observations per se, but the scale of measurement" (Halme et al., 2002, p. 23). By showing the equivalence between considering ratio data or interval data, we demonstrate that in our particular case the use of negative data did not represent a problem. In fact it represented an advantage since we could deal with models having less inputs (if we used ratio data values in period $t-1$ would be on the input side rather on the output side aggregated on variables representing change from one month to the other). Targets, on the other hand, would be less easy to interpret if we have used ratio data because they would imply trade-offs in outputs, but also in inputs due to our weights restrictions. The use of a variations model is, therefore, preferable because it requires less factors to be used in the analysis and it facilitates the interpretation of targets.

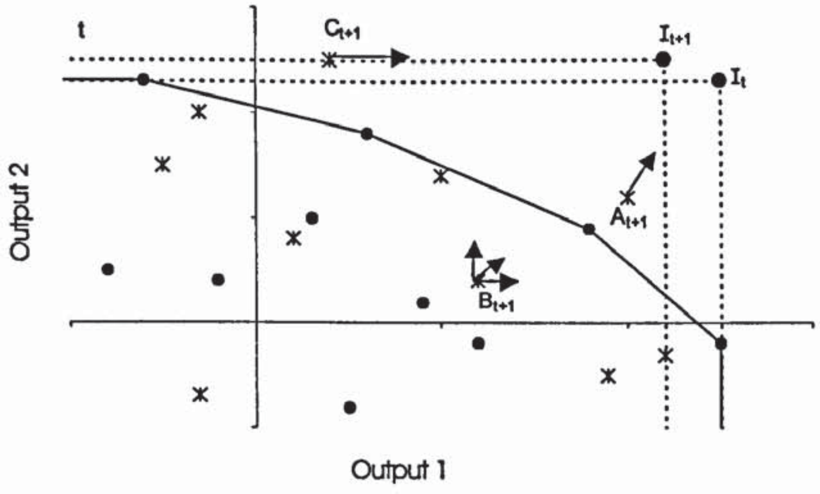
6.6 Malmquist Indexes in the Presence of Negative Data

In the presence of panel data we can assess not only the extent to which efficiency changes, but also the extent to which technology changes over time. Malmquist indexes can be used for this purpose but they have been used so far in the presence of non-negative data only. Since our RDM model provides efficiency scores that are similar in meaning to radial efficiency scores when negative data are present, a natural extension of our approach is to use Malmquist indexes in the presence of negative data.

Using the radial factor $\gamma = 1 - \beta$ resulting from the RDM model we can define a Malmquist index based on this factor as: $MRDM = \left(\frac{\gamma^t(y_{t+1},x_{t+1})}{\gamma^t(y_t,x_t)} \times \frac{\gamma^{t+1}(y_{t+1},x_{t+1})}{\gamma^{t+1}(y_t,x_t)}\right)^{\frac{1}{2}}$, where the superscript in the efficiency measure refers to the frontier against which efficiency is measured. The above MRDM can be decomposed in the usual way into an efficiency change (EFCH) $\left(\frac{\gamma^{t+1}(y_{t+1},x_{t+1})}{\gamma^t(y_t,x_t)}\right)$ factor and into a technological change (THCH) $\left(\frac{\gamma^t(y_{t+1},x_{t+1})}{\gamma^{t+1}(y_{t+1},x_{t+1})} \times \frac{\gamma^t(y_t,x_t)}{\gamma^{t+1}(y_t,x_t)}\right)^{\frac{1}{2}}$ factor. Note that MRDM and its components are defined under VRS since the RDM model in the presence of negative data cannot be used when technologies are CRS. The above THCH is therefore related to the Ray and Desli (1997) decomposition, with the difference that no scale effects will be computed in this case as we cannot calculate MRDM in relation to CRS technologies.

The use of the above Malmquist index implies solving for each period model (6.3) when the frontier is defined by observations from the same period and also when the frontier is defined by observations from a different time period. In the latter case there are some issues that must be accounted for. Take the example shown in Figure 6.4, where we use two outputs to illustrate some problems that might happen. There are units from two time

Figure 6.4: Two Output Example



periods in this figure: units observed in period t (represented with dots) and units observed in $t+1$ (represented with crosses). We will assume that efficiency of both these sets of units is to be assessed in relation to the frontier of period t (the frontier that is drawn in Figure 6.4). Units observed in period t necessarily lie below the t frontier and the value of β lies between zero and one, when the range used is defined in relation to the ideal point of period t represented by I_t in Figure 6.4. As for units observed in period $t+1$, these may lie above or below the t frontier. If we use the $t+1$ ideal point to assess units observed in $t+1$ in relation to the t frontier there is no longer the guarantee that β will be lower than one. As seen previously $\beta = \frac{y^*-y_o}{R_o}$, for some output for which the corresponding constraint is binding. The difference between the target and observed point is lower than the range when all values correspond to the same period. When this is not so, then we have no guarantee that $y^* - y_o \leq R_o$ because the target is no longer on the frontier based on which the range is calculated.

For this reason, the RDM model needs to be modified so that a common ideal point, based on which ranges are calculated, is used in every two periods. This common ideal point is defined as: $Iy_r = \max_t\{\max_j\{y_{rj}^t\}\}$ for outputs and $Ix_i = \min_t\{\min_j\{x_{ij}^t\}\}$. That is, the ideal point is the maximum output (or minimum input) observed in a set of periods that are being considered in the time analysis. For the periods represented in Figure 6.4 the common ideal point is the maximum of output 1 observed in period t and the maximum of output 2 observed in period $t+1$. The range to be used in the RDM model for observations in t and $t+1$ must, therefore, be calculated in relation to this common ideal point. The use of a common ideal point makes all efficiency measures computed in every two time periods comparable and therefore allows the computation of Malmquist indexes. In addition, it makes the vector that departs from the observed point to the ideal point collinear with the vector that departs from the radial target (on the t and $t+1$ frontier) to the ideal point. This collinearity allows the meaningful computation of ratios between the various RDM efficiency measures.

An alternative to the use of a common ideal point in every two periods would be to use a single ideal point corresponding to the maximum output and minimum input levels observed over the entire period of analysis. Although such an alternative would make all efficiency measures comparable, it has the disadvantage of requiring the period of analysis to be fixed *a priori*. In fact, the addition of a new time period to the analysis could imply solving all models backwards again, if the ideal point is changed by this addition.

Furthermore, if the time horizon is very wide the ideal point may be in some months quite distant from observations. As our models are sensitive to the choice of the ideal point it seems preferable to follow a conservative approach where a relative rather than an absolute maximum is considered.

As mentioned above, units observed in a time period that differs from the time period based on which the efficient frontier is drawn may be above or below that frontier. This poses some problems in terms of the feasibility of the RDM model. A unit such as B_{t+1} in Figure 6.4 lies below the t and the $t+1$ frontier and therefore it has always a feasible solution to the RDM model with a positive β. Units that are above the t frontier like A_{t+1}, and C_{t+1}, might present a feasible solution of the RDM, in which case β will be negative (and $\gamma = 1 - \beta$ will be greater than 1), or might have an infeasible solution. For example, unit's C_{t+1} range of improvement leads it to seek to improve only output 1. Following this direction there is no value of β, even negative, that projects unit C_{t+1} on the t frontier. Unit C_{t+1} will therefore have an infeasible solution for the RDM model. Contrary, unit A_{t+1}, which is also above the t frontier has a directional vector that follows approximately the direction shown in Figure 6.4. If we follow this direction in reverse we reach the t frontier, meaning that unit A_{t+1} will have a feasible solution for the RDM model but a negative optimal β. Infeasibility of the RDM model, is always associated with super-efficiency. This means that although we cannot compute the extent to which some units lie above the frontier we know that they lie above it and therefore care is needed in counting the number of units in this situation in order to understand the final average technological change measures.

6.7 Summary

In this Chapter we developed a model that can be used in the presence of negative data. This model is based on the directional distance function approach, where the direction is the range of possible improvement (defined as maximum output minus observed output, or minimum input minus observed input). We call this model Range Directional Model (RDM). The RDM model is units and translation invariant, which makes it suitable to be used in the presence of negative data. In addition, the RDM model results in an efficiency measure that is very similar to those used in radial models except that the point with reference to which efficiency is measured is no longer the origin but an ideal point (having maximum outputs and minimum inputs). Such a measure represents an interesting development in the literature as there was to date no radial or non-radial efficiency measure, to the authors' knowledge, that could be applied directly to negative data.

We extended our approach by considering a variant of the RDM model, where the directional vector is the inverse of the range of possible improvement. The resulting model (IRDM) has the advantage of prioritising improvement of the factors on which the unit performs best, and therefore it tends to yield closer targets to the assessed unit than the RDM model or the well known additive model. The advantage of using both specifications is that production units can choose from different types of targets (one prioritising improvements on the factors on which the unit performs worst, and the other prioritising improvement of the factors on which the unit performs best) both leading to the production frontier.

We also show in this Chapter how the RDM model can be used to compute Malmquist based indexes in the presence of negative data. This is an interesting development introduced here, which was possible due to the radial nature of the RDM efficiency measure. This is, to the authors knowledge, the first attempt in the literature to compute Malmquist indexes when some data are negative.

Chapter 7

Finding Closest Targets[1]

This Chapter presents a novel methodology that requires as low an effort of inefficient production units as possible in their movements towards the production frontier. This is in contrast with some existing methodologies that yield targets for production units that are far away rather than close to their current position. This fact is especially relevant in the context of non-radial and non-oriented efficiency measures as will become clearer throughout the Chapter.

7.1 Introduction

Efficiency measurement in Data Envelopment Analysis (DEA) requires both the identification of a reference point on the boundary of the production possibility set (PPS) and the use of some measure of distance from that point to another being analysed. The two issues (identification of the boundary point and the distance measure used) are traditionally performed simultaneously. The basic DEA model as introduced by Farrell (1957) and later developed by Charnes et al. (1978), uses an oriented radial measure of efficiency, which identifies a point on the boundary with the same mix of inputs (input orientation) or outputs (output orientation) of that of the observed unit. The conservation of the mix in movements towards the boundary of the PPS is the characteristic that makes the resulting distance measure radial.

In many practical situations, however, it is desirable to use measures of efficiency that

[1]Note that part of this Chapter has been published as: Portela, Maria C.A.S., and Borges, Pedro C. and Thanassoulis, Emmanuel (2003), "Finding Closest targets in non-oriented DEA models: the case of convex and non-convex technologies", Journal of Productivity Analysis, 19/2&3.

are non-oriented and non-radial in character. Any measure of efficiency that does not assume equiproportional reductions of inputs or outputs is non-radial. The first non-radial measure of technical efficiency dates back to 1978 and is due to Färe and Lovell (1978). The interest of researchers in non-radial measures arises mainly from the fact that radial (or Farrell) efficiency measures do not necessarily correspond to the Pareto-Koopmans definition of technical efficiency (as already seen in Chapter 2). This issue is known in the DEA literature as the indication or slacks problem, as the main characteristic of radial efficiency measures is that they ignore the possible existence of slacks associated with the projected points on the production frontier. This issue motivated a discussion (e.g. Russell, 1985; Lovell and Schmidt, 1988; Kerstens and Vanden-Eeckaut, 1995) and although many authors still favour the use of radial measures (mainly because of its many useful properties), non-radial efficiency measures have increasing popularity. A weakness of radial measures, is the perceived arbitrariness in imposing targets preserving the mix within inputs or within outputs, when the firm's very reason to change its input/output levels might often be the desire to change that mix (Chambers and Mitchell, 2001, p. 32).

The non-radial Färe-Lovell efficiency measure is oriented. That is, it aims at changing inputs or outputs but not both. To the authors' knowledge, the first (non-radial) non-oriented measures of efficiency were introduced in 1985. One of these, the hyperbolic measure of technical efficiency, is due to Färe et al. (1985) and the other, the additive model, is due to Charnes et al. (1985b). Non-oriented measures are relevant in many practical situations. Take for example the banking context were the use of an intermediation approach (see Colwell and Davis, 1992, for details) specifies inputs in the form of costs and outputs in the form of revenues. Some of the costs and revenues are controllable, and so the obvious approach to follow is non-oriented, i.e, permitting at the same time reduction of inputs and increase of outputs (which in this case would translate in an increase in profits[2]).

The distinction between oriented and non-oriented measures of efficiency is mainly of theoretical interest only, because in practice the analyst needs to identify the variables which can be modified, and then efficiency is measured with reference to those variables. Non-radial oriented measures assume *a priori* that the variables to be modified are only on the input or on the output side, while in practice they are often on both sides. Non-oriented measures are, therefore, more general and more flexible in the sense that they

[2]Profit analysis has been recently advocated in the context of measuring efficiency in banking (see Berger et al., 1993), a field where cost oriented efficiency analysis has been the dominating approach.

allow for changes in all the factors.

One of the key practical outcomes in an efficiency assessment is the identification of targets. However, one of the drawbacks of the traditional non-oriented DEA models is that they either impose strong restrictions on the movements towards the efficient frontier, or they aim at maximising slacks. Both these facts contribute to finding targets and peers that are not the closest to the units being assessed. We will define measures of closeness later; suffice is to say at this point that the closer the targets to a unit, the less the change in its operations needed to reach its targets. If Pareto-efficiency can be achieved by requiring less effort from inefficient units than that demanded by traditional efficiency measures, then it is at least of practical value to find the closest targets for each inefficient unit we can. Close targets in this sense are in line with the original spirit of DEA of showing each production unit in the best possible light.

The idea of finding closest targets and peers has appeared in the literature both associated with oriented models [e.g. Coelli (1998), or Cherchye and Puyenbroeck (2001a)] and non-oriented models [e.g. Frei and Harker (1999), Golany et al. (1993)]. It is our intention to explore this issue for the most general case of non-oriented efficiency measures. In addition, we will restrict our analysis to technical efficiency. In this sense, we shall allow production units to move in all directions to improve their technical efficiency, as long as inputs are not increased and outputs are not decreased.

7.2 Non-Radial-Non-Oriented Measures of Efficiency

Consider a technology represented by $T = \{(\mathbf{x}, \mathbf{y}) \in R_+^{m+s} \mid \mathbf{x} \text{ can produce } \mathbf{y}\}$, where, for each unit j $(1, \ldots, n)$, $\mathbf{x_j} = (x_{1j}, \ldots, x_{mj}) \in R_+^m$ is an input vector producing an output vector $\mathbf{y_j} = (y_{1j}, \ldots, y_{sj}) \in R_+^s$. We address here the two production correspondences $T(\mathbf{x},\mathbf{y})^{FDH}$ and $T(\mathbf{x},\mathbf{y})^{VRS}$, which can both be specified by equation (7.1). When S equals $\{0,1\}$, then T corresponds to an FDH technology ($T(\mathbf{x},\mathbf{y})^{FDH}$) (Deprins et al., 1984), while when S equals $[0, +\infty[$, T corresponds to a VRS (or BCC) technology ($T(\mathbf{x},\mathbf{y})^{VRS}$) (Banker et al., 1984).

$$T(\mathbf{x},\mathbf{y}) = \Big\{(\mathbf{x},\mathbf{y}) \in R_+^{m+s} \mid \sum_{j=1}^{n} \lambda_j \, \mathbf{y_j} \geq \mathbf{y}, \ \sum_{j=1}^{n} \lambda_j \, \mathbf{x_j} \leq \mathbf{x}, \\ \sum_{j=1}^{n} \lambda_j = 1, \ \lambda_j \in S, \ j = 1, \ldots, n\Big\} \quad (7.1)$$

$T(\mathbf{x},\mathbf{y})^{FDH}$ assumes only free disposability of inputs and outputs being, therefore, a non-convex technology. $T(\mathbf{x},\mathbf{y})^{VRS}$, on the other hand, is convex and assumes variable returns to scale. If the constraint normalizing the sum of lambdas was dropped in $T(\mathbf{x},\mathbf{y})^{VRS}$ one would have a constant returns to scale (CRS) technology.

Each one of the two production possibility sets above is bounded by a frontier, where target points are located. A number of measures exist to calculate the distance between observed points and target points. The ones in DEA are known as efficiency measures. Radial measures may find targets that, although lying on the frontier, are not on its Pareto-efficient subset. On the other hand, non-radial measures have the purpose of assuring that the identified targets lie on the Pareto-efficient subset of the frontier. Most of the studies that apply non-radial measures of efficiency use their oriented version [like the Färe-Lovell (see Färe and Lovell, 1978; Färe et al., 1985) or the Zieschang (1984) efficiency measures]. Such studies can be found for example in Dervaux et al. (1998), Ruggiero and Bretschneider (1998), Kerstens and Vanden-Eeckaut (1995), De Borger and Kerstens (1996) or Cherchye and Puyenbroeck (2001b), both in the FDH context and in the context of convex frontiers.

The non-oriented DEA models in the literature share the common feature of maximising slacks. As a consequence, the targets these models identify are the furthest rather than the closest from each production unit being assessed. For some models, like the additive model of Charnes et al. (1985b) or its variant the RAM (Range Adjusted Measure) as proposed by Cooper et al. (1999), this objective of slack maximisation is explicit in the objective function of the DEA models. See for example the objective function of the RAM model that is shown in (7.2), where slacks (normalised by the ranges) are being maximised. The traditional additive model simply maximises the sum of slacks, or alternatively, in one of its units invariant versions, it maximises slacks normalised by the observed input and output levels (see Charnes et al., 1985b; Green et al., 1997).

$$RAM_o = \min\left\{1 - \frac{1}{m+s}\left(\sum_{r=1}^{s}\left(\frac{s_{ro}}{R_r}\right) + \sum_{i=1}^{m}\left(\frac{e_{io}}{R_i}\right)\right)\right\}, \text{where}$$
$$R_r = \max_j\{y_{rj}\} - \min_j\{y_{rj}\}, \quad R_i = \max_j\{x_{ij}\} - \min_j\{x_{ij}\} \quad (7.2)$$

The model of Färe et al. (1985) defined in (7.3) in reference to T (which can be both $T(\mathbf{x},\mathbf{y})^{FDH}$ or $T(\mathbf{x},\mathbf{y})^{VRS}$) also maximises slacks, though this is not explicit in the objective function.

$$FGL_o = \min\left\{\frac{\sum_{i=1}^{m} h_{io} + \sum_{r=1}^{s} 1/g_{ro}}{m+s} \mid (h_{io}x_{io}, g_{ro}y_{ro}) \in T,\ g_{ro} \geq 1,\ 0 \leq h_{io} \leq 1\right\} \quad (7.3)$$

This model is a generalisation of the hyperbolic measure of efficiency[3], where inputs and outputs are allowed to change by different proportions. Using the relationship shown in (7.4) (see Cooper et al., 1999, for details)

$$h_i\, x_{io} = x_{io} - e_i \Leftrightarrow h_i = 1 - \frac{e_i}{x_{io}} \qquad \text{and} \qquad g_r\, y_{ro} = y_{ro} + s_r \Leftrightarrow g_r = 1 + \frac{s_r}{y_{ro}} \quad (7.4)$$

it is possible to show that the objective function of (7.3) is equivalent to:

$$\frac{1}{m+s}\Big(m - \sum_{i=1}^{m}\frac{e_i}{x_{io}} + \sum_{r=1}^{s}\frac{y_{ro}}{y_{ro}+s_r}\Big) \approx 1 - \frac{1}{m+s}\Big(\sum_{i=1}^{m}\frac{e_i}{x_{io}} + \sum_{r=1}^{s}\frac{s_r}{y_{ro}}\Big) \quad (7.5)$$

meaning that model (7.3) also maximises slacks.

The well known directional distance function introduced by Chambers et al. (1996a, 1998) is also a non-oriented measure of efficiency that aims at maximising slacks. Indeed, it is defined as $Dir_o = \max\left\{\beta_o \mid (x_{io} - \beta_o\, g_{x_i},\ y_{ro} + \beta_o\, g_{y_r}) \in T\right\}$, where $\mathrm{g} = (-\mathrm{g_x}, \mathrm{g_y})$ is a directional vector chosen *a priori*. Dividing all inputs and outputs by the directional vector, reduces this measure to the maximisation of a normalised slack value. The directional model is, however, more restrictive than the measures referred to previously in the sense that it strongly limits the direction to be followed towards the production frontier. This means that an optimal solution to Dir_o will potentially result in targets that do not lie on the Pareto-efficient subset of the production frontier, as β cannot account for all the sources of inefficiency. Some references on the use of the above mentioned measures both in FDH and in convex technology settings can be found in De Borger and Kerstens (1996), Bardhan et al. (1996), or Cherchye et al. (2001).

The above mentioned measures will not be used in this Chapter for finding closest targets. Our objective is on the one hand to find an appropriate measure of efficiency and, on the other hand, to operationalise this measure so that closer targets can be found. For reasons that will be explained below, the above measures have some drawbacks in measuring efficiency in a non-oriented context.

[3]Assuming that each h_{io} is constant and equal to θ, and that each g_{ro} is also constant and equal to $\frac{1}{\theta}$ reduces (7.3) to the hyperbolic efficiency measure.

7.2 Non-Radial-Non-Oriented Measures of Efficiency

An appropriate measure of efficiency in a non-oriented context should be capable of incorporating all the sources of inefficiency, while at the same time retaining the meaning of radial efficiency measures. The directional and hyperbolic measures do not satisfy the first requirement, while the RAM, additive model, and FGL model do not satisfy the second requirement. Before showing why this is so, we will present a measure that satisfies both requirements. This is the measure developed by Brockett et al. (1997b), which will be referred to as BRWZ throughout. This measure was originally developed to be used *a posteriori*, that is, after targets have been found, but it can also be used directly in any DEA model. The BRWZ efficiency measure is shown in (7.6).

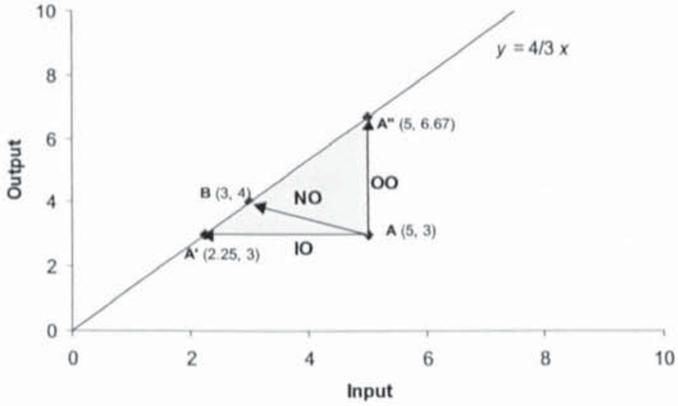
$$BRWZ_o = \frac{1}{m}\Big(\sum_{i=1}^{m} \frac{x_{io} - e_i^*}{x_{io}}\Big) \times \frac{1}{s}\Big(\sum_{r=1}^{s} \frac{y_{ro}}{y_{ro} + s_r^*}\Big) \Leftrightarrow BRWZ_o = \frac{\sum_{i=1}^{m} h_{io} \times \sum_{r=1}^{s} 1/g_{ro}}{m \times s} \tag{7.6}$$

The expression on the left of (7.6) assumes that all inefficiencies are captured by additive slack values (e_i^* and s_r^*, where the star means an optimal value of the input and output slacks as resulting from the solution of some DEA model which projects units on the Pareto-efficient boundary). The equivalent expression on the right of (7.6) (see relationships in (7.4)) makes it possible to show its similarity to oriented measures under certain circumstances[4]. This version of the BRWZ measure is similar to the FGL model in (7.3), but instead of adding the factors on the numerator and denominator it multiplies them. The multiplication of these factors makes the BRWZ measure closer to oriented efficiency measures. To illustrate this fact we present in Figure 7.1 two units in a single input/output space. Unit A is inefficient and it can be projected on the CRS efficient boundary in three different ways. The input oriented efficiency measure (IO) of unit A is 45% which obviously equals its output oriented efficiency measure (OO). Let us assume that the non-oriented (NO) movement of unit A leads to point B. This means that the inputs of unit A should be contracted by 0.6 and outputs expanded by 1.33. With these values the BRWZ efficiency measure equals $0.6 \times \frac{1}{1.33} = 45\%$[5]. The FGL efficiency measure

[4]The BRWZ measure is very similar to the slack based measure (SBM) of Tone (see Tone (1993) and Tone (2001)). The SBM equals $\left(\frac{1}{s}\sum_{r=1}^{s}\left(\frac{y_{ro}+s_r^*}{y_{ro}}\right)\right)^{-1} \times \left(\frac{1}{m}\sum_{i=1}^{m}\left(\frac{x_{io}-e_i^*}{x_{io}}\right)\right)$, which is equivalent to $\left(\frac{\sum_{i=1}^{m} h_i}{m} \times \frac{s}{\sum_{r=1}^{s} g_r}\right)$, when the slacks are replaced by multiplying factors. We prefer the BRWZ to the SBM because the former uses an arithmetic mean of the input efficiency h_i and an arithmetic mean of the output efficiency $1/g_r$. The SBM uses an harmonic mean of the output efficiency whose rationale is not easy to understand.

[5]It is easy to demonstrate that under CRS the BRWZ measure provides the same efficiency measure (45%) for all non-oriented efficient targets in the line segment between A' and A'' in Figure 7.1. Replacing a point

Figure 7.1: Single Input/Output Example



equals $\frac{0.6+\frac{1}{1.33}}{2} = 67.5\%$, and the RAM measure equals $1 - 0.5(\frac{2}{2} + \frac{1}{1}) = 0\%$ if we assume that our sample consists only of A and B, or $1 - 0.5(\frac{2}{2.75} + \frac{1}{3.67}) = 50\%$ if we assume that our sample consists of A, B, A' and A''. If the NO target was point A'', then one would expect non-oriented measures to coincide with the output oriented measure as the projection is the same. This coincidence only happens for the BRWZ efficiency measure which would still be 45%. The FGL efficiency measure would equal $\frac{1+0.45}{2} = 72.5\%$, and the RAM would equal $1 - 0.5(0/2.75 + 3.67/3.67) = 50\%$ or it would be negative if our sample consists only of units A and B.

This simple example shows that the BRWZ measure is indeed closer to the meaning of radial measures as it encompasses as special cases the Farrell radial input and output oriented measures. For example, assuming that all inputs change equiproportionaly (and so each $h_i = \theta$) and that outputs are not allowed to change (and so each $g_r = 1$), the BRWZ measure reduces to θ, which coincides with the Farrell measure of input efficiency. In addition the BRWZ measure is units invariant which is a considerable advantage. The RAM, as Cooper et al. (2001) note, was defined in a VRS technology and should not be applied when CRS prevail. This constitutes an important limitation of this measure. Apart from this, the RAM measure has also the disadvantage of being very sensitive to the composition of the sample as we show above. The inclusion of one unit with an unusually small or large amount of one input (or output) could greatly change the results for many units (see also Steinmann and Zweifel, 2001).

in this line segment by (x, y), and knowing that the line passing trough point A' and A" is $y = \frac{y_{A''}}{x_{A''}}x = \frac{y_{A'}}{x_{A'}}x$, then we have $BRWZ_A = \frac{x}{x_A} \times \frac{y_A}{y} = \frac{x}{x_A} \times \frac{y_A}{\frac{y_{A''}}{x_{A''}}x} = \frac{y_A}{x_A} \times \frac{x_{A''}}{y_{A''}} = \frac{y_A}{y_{A''}}$ because the input at point A and A'' is equal.

One possible disadvantage of the BRWZ efficiency measure is that it weights equally all ratios of target to observed input or output level. Yet not all ratios reflecting short falls from target input-output levels may represent equal 'worth'. This is especially true in contexts where input and output prices are known and shares of inputs and outputs are substantially different between units. While we acknowledge this shortcoming of our measure of distance we do note that it is cast here in the framework of reflecting distance from a technically efficient boundary rather than from some value (cost or revenue) frontier. Distances from value frontiers and associated concepts of allocative efficiency are important but not being addressed here.

7.3 Closer Targets and Efficiency

The objective of finding closest targets implies the definition of closeness. In general, one says that unit B is closer to A than to C, if in order to move from A to B, the changes required in inputs and outputs are smaller than the changes required in order to move from A to C. Such changes can be expressed, for example, in terms of ratios of input and output levels at the two different points concerned. Thus the larger the ratios x^*/x and y/y^*, where the star denotes a target point, the closer the target (x^*, y^*) will be to the unit at (x, y). Obviously in a non-oriented space with multiple inputs and outputs one needs to choose a form of aggregating the above ratios. In our case, the BRWZ efficiency measure was chosen for this aggregation. Thus, the closer the target point to an observed point the higher the BRWZ efficiency as a measure of the distance between the two points.

The closeness between two points can also be measured using an L_p metric. Such metrics are not expressed in ratio form but in difference form. Therefore they have the disadvantage of not being units invariant. The L_p distance between two points (A and B) is given by $\left[\sum_{i=1}^{n} |A_i - B_i|^p\right]^{1/p}$. If A is an observed point and B is a target point on the Pareto-efficient frontier, then L_1 is simply the sum of slacks, as yielded by the additive DEA model. Most of the traditional efficiency measures can be related to L_p metrics as shown by Briec (1998).

We can illustrate concepts of closeness between points using a single input/output example as shown in Figure 7.2. Unit F is FDH and BCC inefficient. In the FDH case unit F is dominated by units B and C. Unit C is closer to F than is unit B. This can be seen in Table 7.1 where the BRWZ measure and some metric distances between points F and C, and F and B are presented. Clearly point B is the point that maximises the sum of slacks

Figure 7.2: FDH Frontier for a Single Input/Output Example

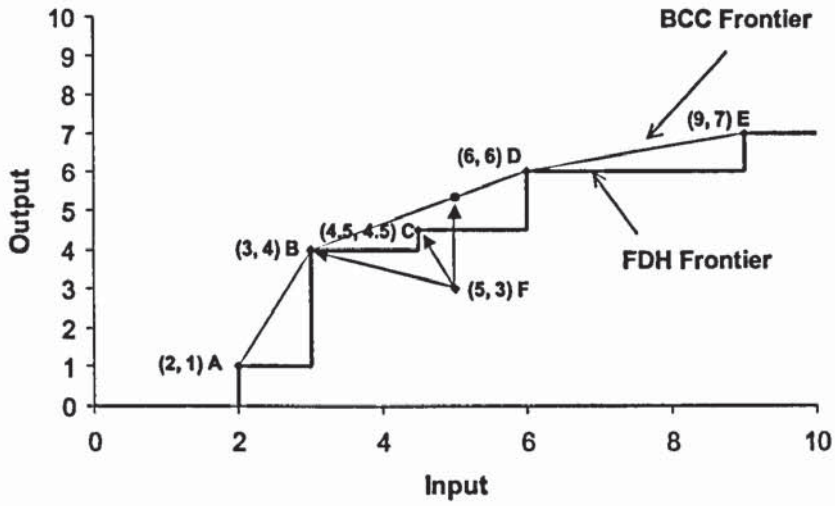


Table 7.1: Distance of F from Points C, B, and (5, 5.33)

Point	$BRWZ$	L_1	L_2	L_∞
B	45%	3	$\sqrt{5}$	2
C	60%	2	$\sqrt{2.5}$	1.5
(5, 5.33)	56.25%	2.33	$\sqrt{5.44}$	2.33

(see L_1 metric), meaning that the non-oriented models mentioned previously - additive, RAM, and FGL - identify point B as the target of unit F rather than point C. This happens both for the case of FDH and convex technology. In the convex context the closest point in terms of the BRWZ measure is point (5, 5.33) - a convex combination between points B and D. Table 7.1 shows that this point is closer to F than the target point B in terms of the BRWZ measure and in terms of the L_1 norm. As far as the other norms are concerned point B seems closer to point F than point (5, 5.33).

We favour the BRWZ measure because it is units invariant, - a characteristic that is important when units of measurement are subjective. The example also shows that traditional DEA models do not necessarily provide targets that are as close as might be possible to the inefficient unit being assessed. As noted earlier, DEA models need to assure that units are projected on the Pareto-efficient frontier and for that purpose maximize slacks. This means that a single stage procedure using the BRWZ incorporated into a DEA structure (see model (7.7)) would not necessarily identify, in the same way as other DEA models, the closest targets.

$$BRWZ_o = \min\left\{\frac{\sum_{i=1}^{m} h_{io} \times \sum_{r=1}^{s} 1/g_{ro}}{m \times s} \mid (h_{io}x_{io}, g_{ro}y_{ro}) \in T\right\} \tag{7.7}$$

Model (7.7) aims at minimising the BRWZ, because this is the only way to assure simultaneously that efficiency is measured and targets lying in the Pareto-efficient frontier are identified. However, these targets are not the closest, as the BRWZ efficiency resulting from (7.7) is not the maximum but the minimum (in the above example the solution of (7.7) identifies unit B as the target of unit F, making the BRWZ measure 45%). To find closest targets one needs to use multi-stage procedures so that we can maximise the objective function in (7.7), while at the same time assuring projection on the efficient frontier. The next two sections will provide means to achieve this, both in FDH technologies and in convex technologies.

7.4 Calculating Closer Targets in FDH Technologies

The interest in finding closer targets in relation to an FDH technology is twofold. First, the targets resulting from efficiency measurement in such a technology correspond to observable units, which might be desirable in some circumstances (for example when inputs and outputs are integer, or when it is likely that the production unit will be more comfortable comparing itself with a real unit rather then with a virtual one - farmers could be such a case). This characteristic makes FDH suitable for benchmarking purposes. Secondly, the non-convex nature of the FDH efficient frontier usually results in higher slack values than those obtained in convex technologies when the direction towards the frontier is restricted in some sense. As noted by De Borger and Kerstens (1996, p. 46) "empirical studies confirm that the amount of unmeasured technical efficiency or slacks is pervasive in FDH". This is a sign that the use of efficiency measures that capture all the sources of inefficiency is potentially more important in FDH than in convex technologies.

The approach developed here takes advantage of the fact that in FDH targets correspond to a single observed unit (peer), which simplifies their identification and the calculation of efficiency. Calculating efficiency requires first the knowledge of the set of dominating units for each dominated unit, and then the selection of the one (the closest) that should be used as the peer.

Our approach follows three steps (note that these steps are usually followed in practical applications of the FDH approach, although not necessarily using the same techniques nor the same criteria for finding peers.):

Step I Determine the set of non-dominated units (100% FDH efficient);

Step II Determine a peer unit for each dominated unit ;

Step III Calculate the efficiency score.

Step I classifies all the units into one of two sets: ND or D. ND is the set of non-dominated (or dominating) units (units in relation to which no other unit exists presenting lower or equal inputs and higher or equal outputs) and D is the set of the remaining units, called dominated. Although this operation can be performed for each unit by comparing it with all the other units or with the current non-dominated set, such implementations become inefficient as the number of units grows. Techniques to handle dominant free sets are also relevant for multiple objective combinatorial optimization, where the state of the art implementations use structures like quad trees for drastically reducing the computational effort spent in such operations (Borges, 2000; Habenicht, 1982). In a quad tree representation of a dominant free set, each node represents a non-dominated unit and can have up to $(2^{s+m}-2)$ branches, which are themselves also quad trees. Each one of those branches corresponds to a particular combination of inputs and outputs in order to guarantee that all units in a branch are dominated by the parent node only in exactly the same outputs and inputs. The discriminatory power of these structures, together with additional bounding techniques, makes them very efficient for handling domination relations and calculating L_∞ metrics. Our implementation uses an algorithm presented in Borges (2000), together with other well known quad tree algorithms to discriminate non-dominated units. These will also be used in step II, to find the units in ND that dominate the unit being assessed. Since these aspects are beyond the focus of the present analysis, whose experiments could just as well have been performed using enumeration, we will not elaborate on them here, for the sake of conciseness.

Step II finds a peer unit for each inefficient or dominated unit. In order to find this unit, we consider a subset of ND, named K, consisting of the units that dominate the unit being assessed. For each inefficient unit, its closest peer is determined through the BRWZ efficiency criterion, that is, calculating (7.8) for every unit $k \in K$.

$$\text{Peer of unit } o = \max_k \left(\frac{\sum_{i=1}^{m} x_{ik}/x_{io}}{m} \times \frac{\sum_{r=1}^{s} y_{ro}/y_{rk}}{s} \right) \tag{7.8}$$

Step III generates the efficiency of the unit being assessed in reference to the peer unit identified in the previous step. The measure of efficiency is given directly from the value obtained in (7.8). Therefore, steps II and III take place simultaneously.

7.5 The Case of Convex Frontiers

Extending the above procedure to convex boundaries is not straightforward because in this case target points may not correspond to observed units but to convex combinations of efficient units. This means that an enumeration oriented procedure which calculates the BRWZ measure for a set of potential target points can no longer be applied. The analogous approach to follow in the case of convex frontiers is to use model (7.7) but with a modified objective function so that the BRWZ is maximised instead of minimised.

Golany et al. (1993) and Frei and Harker (1999) also used DEA models for the case of convex frontiers where the objective function was the minimisation of slacks rather than their maximisation. In the first case, the authors minimised L_1 and L_2 norms in the non-oriented space, and in the second case the authors used L_2 norms for finding the closest targets for each unit, also in the non-oriented space. A concern we share with these authors, is that when distances as the above are being minimised (instead of maximised as is usual in DEA models) it is important to assure that the projected point lies on the Pareto-efficient boundary. For this purpose it is necessary to identify the efficient facets, or at least to have some knowledge about which units belong to which facet. Both, Golany et al. (1993) and Frei and Harker (1999) used the multiplier form of the additive model for the purpose of identifying these facets in the spirit of Ali and Seiford (1993b, p. 130) [see also Huang et al. (1997) and Yu et al. (1996) who used similar procedures for determining efficient facets]. The problem with the use of the additive model is that it does not assure the complete characterisation of the efficient facets. Take the example in Ali and Seiford (1993b, p. 122), where solving the multiplicative additive model for units 1, 4 and 7 (belonging to the same facet) results in 2 hyperplanes: one spanning through units 1 and 4, and the other through units 4 and 7. The full dimension efficient facet (1, 4, 7) is not identified (see also Olesen and Petersen, 1996, who discuss full dimension and non-full dimension efficient facets). This is a direct result of the existence of multiple optimal solutions, which poses a problem in the identification of all efficient facets. Our approach departs from that of Golany et al. (1993) and Frei and Harker (1999) in the procedure used for identifying efficient facets, and also in the distance measure used. The latter is the units invariant BRWZ measure.

There are a number of ways through which efficient facets can be found. In Olesen and Petersen (2002) the authors propose some ways to do so, and in Räty (2002) the author proposes a model that identifies efficient facets (however it is applicable to relatively small problems only) (see also Green et al., 1996, who propose a similar model to that shown in Olesen and Petersen (1996) and Olesen and Petersen (2002)).

For the purpose of identifying efficient facets we use the procedure proposed by Olesen and Petersen (2002), which is based on QHull (see Barber et al., 1996). QHull is a freely available software[6] that can be adapted to the DEA context to identify all full dimension efficient facets (FDEF). Each facet is identified in terms of the convex hull of the Pareto-efficient DMUs whose input-output levels span the facet. The procedure also identifies a supporting hyperplane equation for each facet (for details on the principles behind QHull see Olesen and Petersen, 2002, pp. 32-36). The procedure can also be modified to identify non-full dimensional efficient facets. This involves the use of an augmented data set using artificial DMUs in addition to those observed.

Our procedure for finding the closest targets in convex technologies is divided into three steps:

Step I Determine the set of Pareto-efficient units (E) by solving the additive model;

Step II Determine all Pareto-efficient facets (F_k) using QHull;

Step III For each F_k, $k = 1, \dots, K$ solve model (7.9) to find the closest targets for inefficient unit o.

$$\max\left\{ BRWZ_o = \frac{\sum_{i=1}^{m} h_{io} \times \sum_{r=1}^{s} 1/g_{ro}}{m \times s} \mid \sum_{j\in F_k} \lambda_j\, y_{rj} = g_{ro} y_{ro} \,, \right.$$
$$\left. \sum_{j\in F_k} \lambda_j\, x_{ij} = h_{io} x_{io} \,, \quad \sum_{j\in F_k} \lambda_j = 1 \,, \ \lambda_j \geq 0, \ g_{ro} \geq 1, \ 0 \leq h_{io} \leq 1 \right\} \quad (7.9)$$

In order to assure projection of the efficient frontier only points on F_k are considered as potential projections of unit o in (7.9). The final BRWZ efficiency measure of unit o is the maximum value found for the measure after model (7.9) is solved for all K facets. Step III is repeated for each inefficient unit for which we wish to identify the closest targets. We can also formulate (7.9) as a single mixed integer linear program that should be solved only once for DMU o in respect of all facets K.

[6]http://www.geom.umn.edu/software/qhull

The foregoing procedure has been developed in a VRS context. It is, however, equally applicable to a CRS context. Model (7.9) will change in that the convexity constraint will be dropped. Further, the Pareto-efficient units and the efficient facets will change as we move from a VRS to a CRS technology.

Model (7.9) is non-linear and is not easily linearised. Nevertheless, there are several solvers that can handle non-linear models, whose constraints are linear. We used GAMS and its non-linear programming solver (CONOPT). Nevertheless, for computational convenience a model minimising the sum of normalised slacks, such as that in (7.10), could be used instead.

$$\min \Big\{ \sum_{r=1}^{s} \gamma_{ro} + \sum_{i=1}^{m} \beta_{io} \mid \sum_{j \in F_k} \lambda_j \, y_{rj} = y_{ro} + \gamma_{ro} y_{ro}, \ \sum_{j \in F_k} \lambda_j \, x_{ij} = x_{io} - \beta_{io} x_{io}, \ \sum_{j \in F_k}^{n} \lambda_j = 1, \ \lambda_j \ \geq \ 0, \Big\} \quad (7.10)$$

Such a model, although not equivalent to model (7.9), would likely result in similar targets as a normalised L_1 norm is being minimised. Note that this model is a generalisation of the directional distance function, that assumes a directional vector equal to the observed input and output vectors, and different expansion and contraction factors associated with each input and output. There is one important difference between (7.10) and directional distance function or additive DEA models. The contraction of inputs and expansion of outputs is minimised rather than maximised in (7.10) and this is only made possible by the constraints that ensure the projection point to lie on an efficient facet. The model in (7.10) is also similar to the preference model introduced by Thanassoulis and Dyson (1992). If there are any preferences for moving towards the frontier these can be incorporated in the model in (7.10) as detailed in Thanassoulis and Dyson (1992).

7.6 Illustrative Application

This section applies the above procedures to a sample of 24 Portuguese bank branches which are located in mid sized cities (as classified by the bank) in the northern region of Portugal. This application is simply illustrative of the CT method and should not be confounded with our empirical analysis which is extensively presented in Chapter 10. We use an intermediation approach of banking activities as this requires in principle non-oriented models. In this sense on the input side cost related variables are used (staff costs

Table 7.2: Results from Additive-FDH, RAM-FDH, FGL-FDH and CT Procedure*

Unit	Peer Unit	BRWZ Efficiency	BRWZ CT Efficiency
Unit B3	B10	67.02%	67.02%
Unit B5	B10	77.26%	77.26%
Unit B9	B16	64.70%	64.70%
Unit B13	B10	74.85%	74.85%
Unit B15	B10	53.57%	53.57%
Unit B19	B10	68.15%	81.30% (B20)
Unit B21	B10	71.87%	71.87%
Unit B22	B10	52.76%	78.00% (B52)
Unit B59	B10	74.00%	74.00%

* Results from the Additive-FDH, RAM-FDH, and FGL-FDH are coincident

and other operating costs), and on the output side revenue related variables are used (value of deposit accounts, value of credit, and interest revenues[7]). We assume that all inputs and outputs are discretionary. The data correspond to the month of July 2001 and values are expressed in thousands of Euros. Our input-output set here is only illustrative. Table A.1 in Appendix A contains the data used, as well as some descriptive statistics. The units that were identified as efficient both under FDH and under a convex VRS technology are also identified in this Appendix (see Table A.1). Here we will only detail on the results of some inefficient units.

For the FDH case, the application of the additive units invariant model, the RAM model, and the FGL model result in the same peers for inefficient units in all the cases. This is illustrated in Table 7.2 which shows the BRWZ measure calculated *a posteriori* in relation to the targets identified by these models. It also shows the BRWZ efficiency measure obtained under our closer target (CT) FDH procedure. The BRWZ measure has the same value under all the procedures for identifying targets, except in two cases. The reason for this is simple: unit B10 dominates most of the units in the sample and most of them are solely dominated by this unit. As the set of potential referents consists of a single unit there is not much for the alternative procedures to choose. Only in two cases is there a genuine choice of targets to be made: the case of inefficient units B19 and B22. The first unit is dominated by B10 and also by B20, and the second unit is dominated by B10, B26, B50 and B52. The application of our CT procedure clearly identifies closer targets to units B19 and B22 (respectively B20 and B52) as testifies a higher value of the CT BRWZ

[7]The interest revenues are net of interest costs and this is the reason why these are not considered on the input side. The bank could not provide interest costs and revenues disaggregated.

Table 7.3: Comparison between Models Based on L_p Metrics

—	Additive, RAM, and FGL				CT FDH procedure			
Unit	Peer	L_1	L_2	L_∞	Peer	L_1	L_2	L_∞
Unit B19	Unit B10	4044.94	2920.62	2468.72	Unit B20	800.05	602.98	554.51
Unit B22	Unit B10	6514.76	5367.91	5213.83	Unit B52	1355.58	1004.59	900.77

efficiency score in Table 7.2. These higher efficiency scores also correspond to lower metric distances as can be seen in Table 7.3.

The above example shows that easier-to-achieve targets can be provided to some bank branches, showing them in a better light. If we relied on the traditional models to establish targets we would advise branch B22 to reduce (in thousands of Euros) its staff costs by 6.06 and its other operating costs by 2.96, while at the same time increasing the value of deposit accounts by 1276.8, the value of credit by 5213.83, and its interest revenues by 15.12. Such targets are more demanding than the alternative, which also renders B22 efficient, and corresponds to decreasing staff costs only by 2.02 and operating costs by 3.77, while increasing the value of deposit accounts by 444.73, the value of credit by 900.77, and interest revenues by only 4.29. Only for the case of other operating costs is the target more demanding in this second case, a fact that is more than compensated for by the much less demanding targets in the remaining variables.

In the convex VRS technology case, the application of the CT procedure to the bank branches example results (in its first step) in a set of efficient units that is shown in the last column of Table A.1 in Appendix A. After obtaining the set of efficient units QHull was used to identify the set of efficient facets. These are: F_1 = {B10, B16, B20, B29, B50}; F_2 = {B20, B27, B29, B50, B57}; F_3 = {B10, B20, B27, B29, B50}; F_4 = {B10, B27, B56, B57}; F_5 = {B10, B11, B16, B29}; F_6 = {B10, B11, B26, B29}; F_7 = {B10, B26, B27, B29}, where the first three facets are full dimensional and the last four are not. In the third step, model (7.9) was applied to each inefficient unit in relation to each efficient facet. The facet chosen for projection in each case was the one maximising the objective function of model (7.9). Note that in some cases projection on some facets might be infeasible, but at least one facet shall result in a feasible solution.

The detailed results of applying the additive units invariant model, the RAM model and our CT convex procedure are shown in Appendix A[8]. The results in terms of the various

[8]All results reported concern the use of model (7.9). Model (7.10) was also used and results are equal except for 4 units (B5, B9, B22, B53). In all the cases except B22 the facet of projection was the same both using model (7.9) and model (7.10). Obviously the BRWZ is maximum when model (7.9) is used.

Table 7.4: Distance to Targets for Inefficient Units for the VRS Case

	B15			B59		
	Observed	Targets Additive	Targets CT convex	Observed	Targets Additive	Targets CT convex
x_1	11.717	11.717	11.487	13.338	13.338	12.606
x_2	29.314	24.726	16.122	24.820	24.820	19.030
y_1	4070.630	5682.936	4070.630	4354.301	6073.258	4475.281
y_2	6418.995	14409.226	6418.995	10889.840	14368.013	10889.840
y_3	40.328	69.268	45.086	57.033	74.865	57.033
L_1		9636.066	18.181		5214.962	127.502
L_2		8151.330	14.027		3879.796	121.121
L_∞		7990.231	13.193		3478.173	120.980
BRWZ		53.58%	73.83%		74.56%	84.82%

BRWZ efficiency measures show that *BRWZ CT procedure* $\geq$ *BRWZ Additive* $\geq$ *BRWZ RAM*. We will sidestep the discussion of the relationship between the additive and RAM measures since is not our aim to analyse it here. Concerning the results from our model, they confirm that it shows each inefficient unit in a much better light than the other two models not only in terms of the BRWZ measure but also in terms of L_p metric measures. Take for example units B15 and B59 shown in Table 7.4. Results for these units show closer targets identified by the CT convex procedure than those identified by the additive model (the same being true for the RAM model). This fact is expressed in higher BRWZ efficiency scores and smaller L_p metrics, as illustrated for the two cases above (this fact can however be generalised to the entire sample of units). Interestingly the additive model tends to identify most of the inefficiencies associated with outputs, while the CT procedure for convex technologies identifies most of the inefficiencies associated with inputs. For the additive model the average BRWZ-input efficiency is 98.27% and the average BRWZ-output efficiency is 73.36%, while the corresponding values for the CT procedure are 90.72% and 92.03%, respectively (the RAM model also identifies most of the inefficiencies associated with outputs but to a lesser extent than the additive model: BRWZ-input efficiency is 93.67% and BRWZ-output efficiency is 75.89%). This clearly indicates that our procedure and the additive model identify different directions for improvement of inefficient units. The choice of the model to use should not, thus, be taken lightly.

As a final note on this example one can observe that BRWZ efficiency scores are higher for the convex than for the FDH case, for units that are inefficient under both technologies. The typical result in pure radial models is precisely the reverse because, as it is well known, FDH closely envelops the data and thus provides higher efficiency measures. In our example the closer envelopment resulted in more efficient units for the FDH case but not in higher BRWZ efficiency scores for inefficient units. Note that the range of targets in FDH is limited to observed units, while in the convex case this range is greatly expanded through convex combinations of Pareto-efficient units. This means that we can find closer targets in the convex case than in the FDH case, when we are not restricted to move in any direction and when the measure of efficiency used captures all the sources of inefficiency (that is, when it restricts targets to lie on the Pareto-efficient subset of the frontier).

7.7 Summary

The analysis of non-oriented measures of efficiency and their use to identify the closest targets for inefficient units was performed in this Chapter both considering FDH technologies and convex technologies. The criterion of closeness used is based on the maximisation of the BRWZ measure, which has the advantage over other efficiency measures of capturing all the sources of inefficiency and retaining a meaning that is close to that associated with oriented efficiency measures. In order to use this measure multi-stage procedures are required both in the FDH and in the convex case to find the closest targets. As our analysis restricts targets to lie on the Pareto-efficient subset of the production frontier, the multi-stage FDH procedure starts by choosing potential target units and then it takes the one maximising the BRWZ efficiency measure as the adopted target and peer. In the convex case the aim is also the maximisation of the BRWZ efficiency measure, which results in a non linear programming model, that requires knowledge on the efficient facets of the PPS. The application of our procedure to a real bank branch example shows that it provides closer and easier-to-achieve targets in both, the FDH and convex, cases.

Chapter 8

DEA Models to Measure Profit Efficiency

In this Chapter we introduce a novel approach for measuring profit efficiency in the context of non-parametric methodologies. This is done both considering the long run where all factors of production are variable, and also the short run where some of these factors may be considered fixed.

8.1 Introduction

The Data Envelopment Analysis (DEA) (Charnes et al., 1978) literature has tended to focus on technical efficiency, which can be computed without reference to input or output prices. However, in for-profit organisations technical efficiency alone is of limited interest, as firms will normally be reluctant to change input and output quantities if these do not lead to monetary gains. Obviously the translation of input and output changes into profit requires price information. The often unavailable or inaccurate price information has led, on the one hand, to the widespread measurement of technical efficiency in for-profit organisations pushing to a second place the measurement of economic efficiency, and on the other hand, to the appearance of methodologies that incorporate incomplete price information in efficiency measurement models through the form of weights restrictions (like the cone ratio model of Charnes et al. (1990) and the assurance region model of Thompson et al. (1986)), or through the calculation of inner and outer bounds on efficiency (e.g. Banker and Maindiratta, 1988; Färe and Li, 1998; Kuosmanen and Post, 2001).

Empirical applications measuring profit efficiency have mainly used parametric approaches. Parametric applications to the banking context can be found for example in Humphrey and Pulley (1997), Lozano Vivas (1997), and Berger and Mester (2000). Other applications of parametric profit functions can be found in Lau and Yotopoulos (1971) on agriculture, in Knox et al. (1999) on nursing facilities, and in Kumbhakar (1996) on electric utilities.

Empirical analysis of profit efficiency through the non-parametric DEA approach where input and/or output prices have featured within the assessment model have been rare. Färe et al. (1990) used a profit function to define maximum profit of rice farms, but price data were not considered explicitly and therefore price or allocative inefficiency could not be assessed. Coelli et al. (2002) also used a profit model to determine maximum profit of airline companies, and decomposed a measure of overall profit gap into various components (unused capacity, technical, and allocative inefficiency). However, the authors represented the underlying technology through output distance functions rather than through graph (non-oriented) distance functions as would be consistent with a profit maximising setting. Cherchye and Van Puyenbroeck (2003) also applied output distance functions to German farms, namely the DEA model proposed by Banker et al. (1984), arguing that in the case reliable information on prices is lacking this model provides an estimate of profit efficiency. (The argument used to support this claim is that this model provides 'endogenous benefit-of-the-doubt' prices when price information is lacking).

Another related strand of the DEA literature is the AR (Assurance Region) profit ratios introduced by Thompson and Thrall (1994) and Thompson et al. (1995). As can be seen in Thanassoulis (2001, Chp. 8) DEA weights convey information on marginal rates of substitution/transformation. This interpretation of DEA weights has been used by Thompson et al. (1995) so that the DEA models are redefined to restrict the DEA weights to vary within pre-specified bounds (usually relating to lower and upper bounds on observed prices (e.g. Thompson et al., 1996)). These redefined models are called profit ratios, though they do not in fact measure profit efficiency (e.g. Thanassoulis et al., 2003) as we do in this Chapter. In addition, they differ from the approach presented here, since AR profit ratios do not deal with cases where prices are explicitly used in computing profit.

Profit maximisation implies that at least some inputs and some outputs are endogenous (choice) variables (e.g. Kumbhakar, 2001). If all inputs are exogenously fixed then the maximum profit model reduces to one which seeks maximum revenue. Similarly if all outputs are exogenous then maximum profit would result by minimising cost. If, however, some inputs and some outputs are endogenous then maximum profit can be attained by suitably setting the relative input and output levels to exploit the available input and output prices. The model needed for determining input and output levels that maximise profit would be a 'non-oriented' one in contrast to the traditional oriented models where either inputs are minimised holding outputs relatively stable or the other way round.

As can be gathered from the foregoing in order to measure profit efficiency and lead operating units to improved profitability firstly, a non-oriented approach is needed which will allow for the fact that typically in for profit situations a mix of inputs and outputs is endogenous and not all inputs or all outputs are exogenous. Secondly, while a measure of overall profitability improvement potential is the ultimate aim it is also useful to know whether this can be attained by adjusting the input-output mix to take advantage of prevailing input-output prices or whether profit potential can be exploited by becoming technically more efficient. An overall measure of profit efficiency should, therefore, be decomposed into technical efficiency (entailing information concerning how much a production unit could increase outputs, while at the same time decreasing inputs) and allocative efficiency (entailing information concerning movements along the production frontier from a non-profit maximising point to a profit maximising point, reflecting, as such, failure to maximise profits under prevailing prices). Existing approaches on the literature that tackle these twin issues of non-orientation and efficiency decomposition (e.g. the hyperbolic model of Färe et al. (1985) or the directional model of Chambers et al. (1998)) have some flaws that we try to overcome in this Chapter. These flaws relate to the fact that these models do not in general account for all sources of inefficiency, and also to the fact that they may calculate profit efficiency on the basis of points located outside the production possibilities set.

8.2 Measuring Profit Efficiency

Consider a technology represented by $T = \{(\mathbf{x}, \mathbf{y}) \in R_+^{m+s} |\ \mathbf{x} \text{ can produce } \mathbf{y}\}$, where $\mathbf{x}$ is an input vector ($i = 1, ..., m$), and $\mathbf{y}$ is an output vector ($r = 1, ..., s$). We will assume throughout that T is represented by the following production possibility set:

$$T = \Big\{(\mathbf{x},\ \mathbf{y})\ \in\ R_+^{m+s} |\ \sum_{j=1}^{n} \lambda_j\, \mathbf{y_j} \geq \mathbf{y},\ \sum_{j=1}^{n} \lambda_j\, \mathbf{x_j} \leq \mathbf{x},\ \sum_{j=1}^{n} \lambda_j = 1,$$
$$\lambda_j \geq 0,\ j = 1, \ldots, n\Big\} \quad (8.1)$$

That is, T is a monotonic (or strongly free disposable) convex set satisfying variable returns to scale (VRS) (for details see e.g. Färe et al., 1994a).

Within the DEA framework, whatever the approach used to calculate profit efficiency, the starting point of a profit analysis is the calculation of maximum attainable profit. This can be done using the model shown in (8.2) (e.g. Färe et al., 1994a, p. 213).

$$\max_{\lambda_j, y_r, x_i} \left\{ \sum_{r=1}^{s} p_{ro}\, y_r - \sum_{i=1}^{m} w_{io}\, x_i \mid \sum_{j=1}^{n} \lambda_j\, y_{rj} - y_r \geq 0 \;\; r = 1, \ldots, s \,, \right.$$
$$\left. \sum_{j=1}^{n} \lambda_j\, x_{ij} - x_i \leq 0 \;\; i = 1, \ldots, m \,, \; \sum_{j=1}^{n} \lambda_j = 1 \,, \; \lambda_j \geq 0 \;\; , j = 1, \ldots, n \right\} \quad (8.2)$$

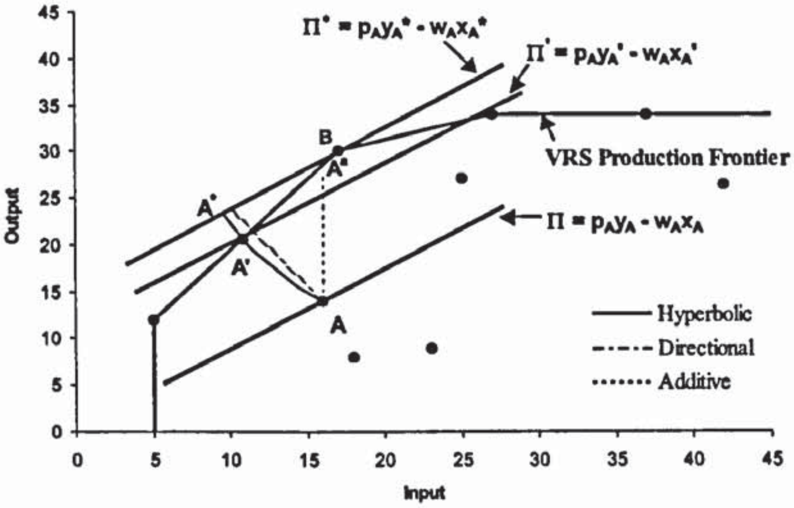
Model (8.2) assures profit maximisation in the long run as no factors are considered fixed. Furthermore, it considers no other constraints apart from technological (though see Färe et al., 1990, where expenditure constraints were added to (8.2)). Model (8.2) assumes VRS since for a technology exhibiting globally constant returns to scale (CRS) either the maximum profit level is zero or the solution of the maximum profit model is undefined (e.g. Varian, 1992; Färe et al., 1994a).

The implications of assuming VRS in (8.2) are: (i) We do not assume perfectly competitive markets since under this assumption all firms have zero profits in the long run. In (8.2) maximum profit may be positive. It should be noted that the concept of profit that economic theory deals with is **economic profit** - i.e., the profit that includes all factors of production evaluated at their opportunity cost (e.g. Varian, 1992, p. 23). We do not maintain the assumption of perfectly competitive markets as we cannot assess economic profit at opportunity cost. (ii) Scale efficiency cannot be calculated as a component of overall profit efficiency. In order to make this possible the maximum profit model (8.2) should be applied under CRS (e.g. Färe et al., 1994a). (iii) Maximum profit units do not need to be most productive scale size (mpss) units in the sense of Banker (1984). That is, maximum profit units do not need to be scale efficient (see also Kuosmanen, 1999b). The implications resulting from a VRS assumption will be further addressed throughout this Chapter.

8.2.1 Brief Review of Existing Approaches

In this section we briefly review existing approaches to measure and decompose profit efficiency by means of the technically and profit inefficient unit A in Figure 8.1. Unit A

Figure 8.1: Profit Efficiency Measurement



achieves maximum profit when it is projected on the profit frontier (say at A^*), where maximum profit equals that of unit B, a maximum profit unit. If *overall profit efficiency* is measured by means of a ratio between profit at two points, then overall profit efficiency of unit A is given the ratio $\frac{\Pi}{\Pi^*}$ (see e.g. Banker and Maindiratta, 1988; Cooper et al., 2000, who used this ratio-based approach within the DEA framework). The *technical profit efficiency* of this unit can also be calculated by a ratio of profits, namely as $\frac{\Pi}{\Pi'}$, where Π' is the profit at the technically efficient projection A' of A. The *allocative profit efficiency* of A ($\frac{\Pi'}{\Pi^*}$) can now be calculated by decomposition from $\frac{\Pi}{\Pi^*} = \frac{\Pi}{\Pi'} \times \frac{\Pi'}{\Pi^*}$.

The aforementioned ratio-based approach is analogous to what is usually done in cost or revenue settings where ratios of cost or revenue represent efficiency (see Chapter 2 section 2.3.1). In such settings, however, there is no possibility of negative costs or revenues. This is not the case with profits that can be negative. The problem of negative profits was not recognised as such by some authors. E.g. Berger and Mester (2000, p. 98) state that "profit efficiency can be negative, since firms can throw away more than 100% of their potential profits". Others like Banker and Maindiratta (1988) assume that all production units exhibit positive profit. Finally some authors have acknowledged this problem and solved it either by using revenue/cost ratios, which can never be negative[1] (see e.g Cooper et al., 2000; Kuosmanen, 1999b) or by using differences between profits to avoid negative

[1]Under this approach the profit efficiency of unit A (see Figure 8.1) would be defined as a ratio of revenue-to-cost so that: $\frac{p_A y_A / w_A x_A}{p_A y_A^{**} / w_A x_A^{**}}$, where the double star is the optimal solution of: $\max\{\frac{p_A y}{w_A x} | (x, y) \in T\}$ (see for details Cooper et al., 2000).

efficiency measures (e.g. Berger et al., 1993; Coelli et al., 2002).

Rather than using the above ratio-based approach some authors defined overall profit efficiency as being a measure reflecting the required adjustments on the input/output levels of unit A that moves it to point A^* on the profit frontier. For example, the **hyperbolic model** of Färe et al. (1985) (see also Färe et al., 1994a) defines the technical profit efficiency of unit A as being θ_A defined in $\min\{\theta_A \mid (\mathbf{x_A}\theta_A \ , \ \mathbf{y_A}/\theta_A) \ \in \ T\}$. In accordance with this hyperbolic path the overall profit efficiency (ϕ_A^h) of unit A, is derived by solving $\Pi^* = \mathbf{p_A}(\mathbf{y_A}/\phi_A^h) - \mathbf{w_A}\mathbf{x_A}\phi_A^h$ where $[\Pi^*]$ is the maximum profit of A calculated through model (8.2). That is, overall profit efficiency (ϕ_A^h) represents the amount by which inputs and outputs should be hyperbolicly adjusted, so that they are projected on the profit boundary[2]. The overall profit efficiency can then be decomposed as: $\phi_A^h = \theta_A \times \gamma_A^h$, where γ_A^h is the allocative profit efficiency.

The **directional model** of Chambers et al. (1996a, 1998) follows a procedure that is similar to that of the hyperbolic model, except that the overall profit efficiency (ϕ_A^d), would decompose as: $\phi_A^d = \beta_A + \gamma_A^d$, where β_A represents technical profit inefficiency and γ_A^d represents allocative profit inefficiency (for details see Chambers et al., 1998).

When the **additive model** of Charnes et al. (1985b) is used, then a technically efficient target (A^a in Figure 8.1, whose profit is Π^a), and a maximum profit target (B in Figure 8.1, whose profit is Π^*) are identified. Using these targets Cooper et al. (1999, 2000) decomposed the profit lost due to overall profit inefficiency into the profit lost due to technical and allocative inefficiency (i.e. $(\Pi^* - \Pi) = (\Pi^a - \Pi) + (\Pi^* - \Pi^a)$). This relationship is not, however, expressed in efficiency terms but in absolute profit values (see Berger et al., 1993; Coelli et al., 2002, who also used profit differences, though not using the additive model).

The foregoing approaches have certain disadvantages in a practical context which the profit efficiency measure we put forward in this Chapter overcomes. The key drawbacks of the foregoing measures are as follows.

Negative efficiency measures. It is possible for the ratio-based profit efficiency approaches to result in a negative measure of efficiency when profit is negative (a loss). Such measures generally cannot be interpreted in a practical way to reflect distance from maximum profit and are therefore problematic. Efficiency measures which take only positive values are to be preferred. The measure of efficiency we put forward here can only take

[2]When the technology satisfies constant returns to scale maximum profit is assumed to be zero and the above expression simplifies to $\phi_A^h = \sqrt{\frac{\mathbf{p_A}\mathbf{y_A}}{\mathbf{w_A}\mathbf{x_A}}}$ (see Färe et al., 1985, 1994a, for details).

positive values.

Reference to Infeasible Points. The hyperbolic and directional models may calculate overall profit efficiency with reference to infeasible points such as A^* in Figure 8.1. This can also happen in 'oriented' cost or revenue settings, but in these cases projections on infeasible points can be interpreted in terms of ratios of inputs (outputs) between the observed and the infeasible point because such ratios match the ratio of minimum cost (maximum revenue) to that at the observed point. This is no longer so in the non-oriented profit setting. For example, if we assume a hyperbolic path is followed from A to A^* in Figure 8.1, then the required adjustments in inputs and outputs are given by ϕ_A, as $x_{A^*}/x_A = y_A/y_{A^*} = \phi_A$. The profit ratio, on the other hand equals $\frac{\Pi}{\Pi^*} = \frac{p_A y_A - w_A x_A}{p_A y_{A^*} - w_A x_{A^*}} = \frac{p_A y_A - w_A x_A}{p_A y_A/\phi_A - w_A x_A \phi_A}$, which differs from ϕ_A. (Note that this statement is valid for all paths).

Accounting for slacks. The calculation of technical profit efficiency through the hyperbolic and directional models assumes the same factor (θ or β) associated simultaneously with inputs and outputs. The resulting efficiency measures do not account, therefore, for all the sources of inefficiency, namely those associated with slacks. This is an important problem in a context where overall efficiency is being measured because what is not captured by technical efficiency will be incorporated into allocative efficiency, which may therefore be incorrectly estimated. The measure of efficiency we put forward in this Chapter accounts for all sources of inefficiency.

Best alternative targets for technical efficiency. The additive model does not have the above problems, but it does not yield an efficiency score. In addition, it provides technical efficient targets to inefficient units that may be remote targets rather than being as close as possible to the observed unit (e.g. Portela et al., 2003). Note that Chavas and Cox (1999) also showed the concern for closest technical efficient targets in a model where the authors measure and decompose profit efficiency (we do not detail their model in this Chapter for sake of brevity). In our approach we share this concern, and calculate technically efficient targets that are the easiest for the firm to attain.

8.3 A Geometric Distance Function

We develop here a new measure of profit efficiency. It is based on a geometric mean measure of distance between the point whose efficiency is being measured and the maximum profit point. We give the definition and the properties of the measure before discussing its advantages over existing measures of profit efficiency.

Radial measures of efficiency can be expressed as ratios between productivity measures. Consider in a two-dimensional space an observed point (x, y) and a target point (x^*, y^*) on the frontier of T, where T is defined in (8.1). The radial output efficiency measure equals $\frac{y}{y^*}$, and the radial input efficiency measure equals $\frac{x^*}{x}$. Both measures are in fact equal to a ratio between the productivity at the observed and at the target point $(\frac{y/x}{y^*/x^*})$. The output efficiency measure is derived if we set in the foregoing ratio $(x^* = x)$ and the input efficiency if we set $(y^* = y)$. Note now that $(\frac{y/x}{y^*/x^*})$ can also be expressed as a ratio of target to observed input and output levels: $\frac{x^*/x}{y^*/y}$. We generalise this latter ratio to a multi-input multi-output context as follows.

Consider in the multiple input/output space an observed vector $(\mathbf{x}, \mathbf{y})$ and a target vector $(\mathbf{x}^*, \mathbf{y}^*)$ on the frontier of T. The traditional radial approach for measuring technical efficiency takes the maximum of $\frac{\mathbf{x}^*}{\mathbf{x}} = \max(\frac{x_1^*}{x_1}, \frac{x_2^*}{x_2}, \ldots, \frac{x_m^*}{x_m})$ (which equals $\max(\frac{\mathbf{x}^*/\mathbf{x}}{\mathbf{y}^*/\mathbf{y}}) = \frac{\max(\mathbf{x}^*/\mathbf{x})}{\min(\mathbf{y}^*/\mathbf{y})}$ when $\mathbf{y} = \mathbf{y}^*$) as the input radial efficiency measure. Similarly, the maximum of $\frac{\mathbf{y}}{\mathbf{y}^*}$ or the minimum of $\frac{\mathbf{y}^*}{\mathbf{y}}$ (which equals $\max(\frac{\mathbf{x}^*/\mathbf{x}}{\mathbf{y}^*/\mathbf{y}})$ when $\mathbf{x} = \mathbf{x}^*$) is the output radial efficiency measure. This means that the analogous situation in the non-oriented case would be to take the maximum of $\frac{\mathbf{x}^*/\mathbf{x}}{\mathbf{y}^*/\mathbf{y}}$ as the efficiency measure, where no inputs or outputs are considered fixed. If we now assume that between $(\mathbf{x}, \mathbf{y})$ and $(\mathbf{x}^*, \mathbf{y}^*)$ all inputs change by the same proportion (say $\mathbf{x}^* = \theta\mathbf{x}$), and all outputs change by the same proportion (say $\mathbf{y}^* = \beta\mathbf{y}$), then $\frac{\mathbf{x}^*/\mathbf{x}}{\mathbf{y}^*/\mathbf{y}} = \frac{\theta\mathbf{x}/\mathbf{x}}{\beta\mathbf{y}/\mathbf{y}} = \theta/\beta$ corresponds to the non-oriented efficiency measure, defined as a ratio of target to observed input level divided by the ratio of target to observed output level. If finally we consider the more general case where inputs and outputs are allowed to change by different proportions (so that $x_i^* = \theta_i x_i$ and $y_r^* = \beta_r y_r$) then the measure in (8.3) parallels the foregoing ratio of target to observed input-output levels.

$$\text{Geometric Distance Function (GDF)} = \frac{(\Pi_i \theta_i)^{1/m}}{(\Pi_r \beta_r)^{1/s}} \tag{8.3}$$

It is this measure that we propose to use to measure profit efficiency.

The measure in (8.3) has been named geometric distance function (GDF), because it is the ratio of the geometric mean distance between observed and target input levels and the corresponding distance between observed and target output levels. The geometric distance function can be used to measure the distance between any two points. When these points relate to maximum profit and current point, respectively, the resulting efficiency is *overall profit efficiency.* If they relate to a technically efficient point and the current point, the resulting GDF is a general measure of *technical efficiency.*

The geometric distance function incorporates as special cases the usual input and output oriented measures of efficiency in DEA. Consider a GDF technical efficiency measure of $(\mathbf{x_o}, \mathbf{y_o})$ as the solution of model (8.4).

$$GDF(\mathbf{x_o}, \mathbf{y_o}) = \min \left\{ \frac{(\Pi_i \theta_{io})^{1/m}}{(\Pi_r \beta_{ro})^{1/s}} \mid \sum_{j=1}^{n} \lambda_j \, y_{rj} = \beta_{ro} y_{ro}, \; \sum_{j=1}^{n} \lambda_j \, x_{ij} = \theta_{io} x_{io}, \right.$$
$$\left. \sum_{j=1}^{n} \lambda_j = 1, \; \lambda_j \geq 0, \; 0 \leq \theta_{io} \leq 1, \; \beta_{ro} \geq 1 \right\} \quad (8.4)$$

Model (8.4) is highly non-linear, but it becomes linear for its special cases of traditional oriented models. For example, in input oriented measures one assumes that $\beta_{ro} = \beta = 1 \, \forall \, r$, and also that $\theta_{io} = \theta \, \forall \, i$, and thus the final efficiency score in (8.4) reduces to θ, which corresponds to the Farrell input efficiency measure. For output oriented efficiency measures similar reasoning applies. If $\beta_{ro} = \beta \, \forall \, r$ is assumed to equal the inverse of $\theta_{io} = \theta \, \forall \, i$, then (8.4) reduces to the hyperbolic model of Färe et al. (1985), with the only difference being that the resulting measure of efficiency is θ^2 and not θ.

The advantage of allowing for different contraction factors associated with inputs (θ_i) and different expansion factors associated with outputs (β_r) is that all sources of inefficiency can be captured by the resulting efficiency measure. This is not a new concern. For example, Färe et al. (1985) introduced a 'Russell graph measure of technical efficiency' that allows for different changes within inputs and outputs. However, the authors define the arithmetic average of the various factors associated with inputs and outputs, $\left(\frac{\sum_{i=1}^{m} \theta_i + \sum_{r=1}^{s} 1/\beta_r}{m+s}\right)$, as the final efficiency measure (see also Pastor et al., 1999, who propose an 'enhanced Russell graph efficiency' measure). We argue that in order to retain the meaning of traditional efficiency measurement the geometric average between these factors should be used instead. In addition, the use of the geometric mean has the advantage of allowing decomposition of an overall measure of efficiency into its technical and allocative components, which is a characteristic that the Russell Graph measure does not possess (this is also the case for other measures in the literature that try to include all the sources of inefficiency such as the SBM (slacks-based measure) of Tone (2001) or the RAM model of Cooper et al. (1999)). However, when we use the GDF measure, if we consider an observed point $(\mathbf{x}, \mathbf{y})$, a maximum profit point $(\mathbf{x}^*, \mathbf{y}^*)$, and a technically efficient point $(\mathbf{x}', \mathbf{y}')$, then a profit measure can be decomposed as shown in (8.5).

$$\text{Overall Profit Efficiency} = \frac{(\Pi_i \frac{x_i^*}{x_i})^{1/m}}{(\Pi_r \frac{y_r^*}{y_r})^{1/s}} = \frac{(\Pi_i \frac{x_i'}{x_i})^{1/m}}{(\Pi_r \frac{y_r'}{y_r})^{1/s}} \times \frac{(\Pi_i \frac{x_i^*}{x_i'})^{1/m}}{(\Pi_r \frac{y_r^*}{y_r'})^{1/s}} \tag{8.5}$$

That is, **Overall Profit Eff. = Technical Profit Eff. × Allocative Profit Eff.** We shall return to this point in section 8.3.2 when we calculate each of these components.

8.3.1 Properties of GDF-based measures of profit efficiency

We use model (8.4) to highlight some general properties of the GDF measure.

Intuitive Homogeneity - The GDF technical efficiency measure in (8.4) is sub-homogeneous of degree -2 (i.e., when inputs are halved and outputs are doubled the measure of efficiency increases by a factor of at least 4 times)(see appendix B for proof). To the authors' knowledge there is no non-oriented efficiency measure in the literature that satisfies homogeneity of degree -2. When we look just at one side (inputs or outputs) it seems intuitive to say that if all outputs double the measure of efficiency should double (homogeneity of degree -1). When both inputs and outputs are being changed, homogeneity of degree -1 is not intuitive as a simultaneous change in both inputs and outputs should bring about a larger change in efficiency than when only inputs or outputs are changed. Nevertheless, most of the existing non-oriented measures in the literature satisfy degree of homogeneity -1, like for example the hyperbolic measure or the Russell graph measure of technical efficiency (Färe et al., 1985).

Monotonicity - The GDF technical efficiency measure in (8.4) is weakly monotonous on inputs and on outputs. That is, when inputs increase and/or outputs decrease the measure cannot improve (see appendix B for proof).

Reflecting all inefficiencies - When the GDF is used to measure technical profit efficiency, then it is a measure incorporating all the sources of inefficiency. Indeed, note that we assumed equalities in the constraints of model (8.4), which results, by definition, in zero slacks. This means that the GDF-based measure of technical profit efficiency is 1 (100%) if and only if the unit being assessed is Pareto-efficient.

Lower and Upper bounds - The GDF technical profit efficiency in (8.4) varies between 0 and 1 (see appendix B for proof). In contrast the allocative profit efficiency and the overall profit efficiency measure can take any positive value. The measure takes the value 1 when the two points (e.g. maximum profit point and observed point) used are coincident. However, the converse is not true and the GDF overall profit efficiency can be 1 even when the two points used in the measure are not coincident. For example if one input is halved and another is doubled the result in the numerator of the GDF is 1, which does not mean that inputs did not change but that on average they stayed the same. Because of the averaging process within the computation of the overall profit efficiency it is necessary to further decompose it, so that one can understand and interpret its value. In the next section we will show how each component of (8.5) is calculated and interpreted.

Economic interpretation - It is interesting to note that the GDF measure of profit efficiency can also be expressed in terms of the ratio of the geometric mean of cost and revenue changes as we move from a current (i.e. observed) to a maximum profit point. In particular note that we can replace $\frac{(\Pi_i \frac{x_i^*}{x_i})^{1/m}}{(\Pi_r \frac{y_r^*}{y_r})^{1/s}}$, by $\frac{(\Pi_i \frac{w_i x_i^*}{w_i x_i})^{1/m}}{(\Pi_r \frac{p_r y_r^*}{p_r y_r})^{1/s}}$ which equals $\frac{(\Pi_i \frac{C_i^*}{C_i})^{1/m}}{(\Pi_r \frac{R_r^*}{R_r})^{1/s}}$. Thus we can see the numerator (and denominator) of the GDF either as the geometric average of changes in input quantities (and output quantities) or as the geometric average of changes in costs (and revenues) resulting from moving from actual to maximum profit.

As noted earlier, in radial input efficiency measures we assume fixed output levels and all inputs are contracted by the same factor. In such a case the GDF measure reduces to the ratio of minimum attainable cost to observed cost, precisely the traditional measure of overall cost efficiency. A radial cost efficiency expresses, therefore, both the required radial reduction in inputs and the percentage cost savings achieved by this reduction. Unfortunately, when we are in non-oriented space an analogous interpretation of the GDF efficiency in terms of profit is not possible. Even so, the GDF shows a link between quantities and profit in the sense that to increase profit the average change in revenues should, in principle, be more than proportional to the average change in costs.

8.3.2 The GDF for Measuring and Decomposing Profit Efficiency

Having defined the profit efficiency measure in the previous section we discuss here the practical steps involved in its calculation and decomposition.

For calculating overall profit efficiency we shall first use model (8.2) and then apply the geometric distance function in (8.3) to measure the 'distance' between observed points and maximum profit points. The resulting GDF measure is the overall profit efficiency.

As noted earlier the GDF overall profit efficiency can take any positive value. A value of 1 is a necessary but not sufficient condition for the point to be maximum profit. A GDF overall profit efficiency equal to 1 indicates that the geometric average change in inputs equals the geometric average change in outputs, which might happen when the two points being compared are the same or not. Therefore, the necessary and sufficient condition for overall profit efficiency is that the maximum profit point is coincident with the observed point in which case the overall profit efficiency measured by GDF will be 1. Any other value of the GDF overall profit efficiency will indicate a profit-inefficient point. A value above 1 indicates that the geometric average change in inputs is higher than the geometric average change in outputs, and a value below 1 indicates that the geometric average change in inputs is lower than the geometric average change in outputs. So the overall profit efficiency measure conveys information on how much we have to change the physical value of inputs and outputs in order to improve profit.

The decomposition of overall profit efficiency that we propose here is carried out in two phases. In phase 1 we decompose overall profit efficiency and in phase 2 we decompose allocative profit efficiency into a scale and a pure mix effects.

Phase 1 - Decomposing the GDF overall profit efficiency

Recall that overall profit efficiency can be decomposed as: Technical profit efficiency × Allocative profit efficiency. The calculation of (**technical profit efficiency**) could be done using the geometric distance function in (8.4). However, this model is highly non-linear and therefore we propose to measure technical profit efficiency *a posteriori* after a technically efficient target has been identified.

Procedures for finding such targets may be any one of the existing (non-oriented) in the literature. We shall, however, depart from these for two reasons. For example we do not wish to strongly restrict movements of production units towards the production frontier. We do not see any reason why the mix within inputs and within outputs should be preserved as imposed by the hyperbolic or directional models. The motive behind these assumptions seems to have to do more with mathematical convenience, while in reality "the firm's very reason to change its output vector may be to change the mix" (Chambers and Mitchell, 2001, p. 32). The only restriction that we impose is that inputs cannot increase and outputs cannot decrease in moving towards the production frontier. However, note that this restriction only applies in technical efficiency measurement and it is consistent with the notion of looking for input-output improvements which may be feasible before any trade-offs consequent on input and/or output prices are taken into account. Secondly, we want to find the closest targets on the production frontier (and not the farthest (using the

L_1 norm criterion) as the additive model does[3]. The reason for selecting closest targets relate to the fact that they are easier in practice to achieve, and more in line with the way management exercise judgment in general.

Here we follow a similar approach to that introduced in Chapter 7 to calculate the closest technically efficient targets based on the minimisation of normalised slacks. The procedure to be used here consists of the following four steps:

- Apply the additive model under VRS to find the set of Pareto-efficient units (E);
- Find all the Pareto-efficient facets ($F_k, k = 1, \ldots, K$) of the production frontier;
- Find closest targets to the unit at hand on the Pareto-efficient frontier;
- Calculate for the unit at hand a technical efficiency measure using the GDF in (8.3).

The first step is straightforward and does not require further comment. The second step is potentially the hardest to perform but there is software that can easily handle the problem. Olesen and Petersen (2002) have used QHull (a freely available software[4]) for finding all 'Full Dimensional Efficient Facets' (FDEF) in a DEA model (see also Olesen and Petersen, 1996). The required input for this software is the input/output levels of efficient units identified in the previous step, based on which Qhull identifies all the units that lie on each FDEF, and provides a supporting hyperplane for each facet. QHull can also be used to identify non-full dimension efficient facets (NFDEF), but in this case its application is not straightforward. In this Chapter we only use FDEFs, though the use of NFDEFs was illustrated in Portela et al. (2003).

After finding all the efficient facets $F_k,\ k = 1, \ldots, K$ through Qhull, we are able to apply model (8.6), that identifies the closest targets in terms of minimum sum of normalised slacks, for each inefficient unit relative to the FDEF (F_k).

$$\begin{aligned}\mathrm{EFF}_o = \min \Bigg\{ \sum_{r=1}^{s} \gamma_{ro} + \sum_{i=1}^{m} \beta_{io} \mid \sum_{j \in F_k} \lambda_j\, y_{rj} = y_{ro} + \gamma_{ro} y_{ro} \\ \sum_{j \in F_k} \lambda_j\, x_{ij} = x_{io} - \beta_{io} x_{io}\,,\ \sum_{j \in F_k}^{n} \lambda_j = 1,\ \ \lambda_j,\ \gamma_{ro},\ \beta_{io}\ \geq\ 0 \Bigg\}\end{aligned} \tag{8.6}$$

[3]The geometric distance calculated through model (8.4) does not accomplish the closest targets objective either. Note that the GDF in (8.4) can be expressed as a function of slacks by replacing each θ_i by $1-(e_i/x_{io})$ and each β_r by $1+(s_r/y_{ro})$ (see Cooper et al., 1999, where these relationships are put forward). The GDF is thus equivalent to a slacks' based measure, where slacks are being maximised rather than minimised.

[4]www.geom.umn.edu/software/qhull

Model (8.6) is in fact equivalent to an additive units invariant model where normalised slacks are being minimised instead of maximised[5]. Model (8.6) must be solved for each inefficient unit in relation to each efficient facet identified in the previous step. Obviously this model may, in some situations, be infeasible (if the efficient facet being considered is not attainable by the inefficient unit being assessed) but there will be at least one facet for which the above problem has a solution. The facet to be chosen for projection is the one corresponding to the lowest optimal objective function in (8.6). The targets resulting from this projection can then be used to calculate the technical efficiency measure (4^{th} step).

Given the above, we define technical profit efficiency as a distance between the observed point and the technical efficient point that is its closest target. In moving from the observed to the technical efficient point a unit may change the mix of inputs and/or outputs but these changes in mix are not sought to in any way reflect factor prices.

In terms of interpretation, as the technical profit efficiency component ranges from 0 to 1 (see the appendix B), a value of 1 means the observed point is Pareto-efficient, while a value below 1 means the observed point is not Pareto-efficient.

Note that the way technical efficiency is measured has obvious implications on the amount of overall inefficiency that is explained by technical and by allocative sources. By choosing closest targets we are 'minimising' the component attributable to technical inefficiency and 'maximising' the component that is attributable to allocative inefficiency.

Having calculated the overall profit efficiency and the technical profit efficiency, the **allocative profit efficiency** can be calculated as the ratio of overall profit efficiency and technical profit efficiency (see decomposition in (8.5)). The allocative profit efficiency reflects movements from a technically efficient point $(\mathbf{x}', \mathbf{y}')$ to a maximum profit point $(\mathbf{x}^*, \mathbf{y}^*)$. Such movements imply changes in the mix of inputs and/or outputs that are dictated by factors prices. However, movements from a technical efficient point to a maximum profit point may not only imply changes in mix but also changes in scale size. As noted by Lovell and Sickles (1983, p. 54) for a profit setting, "in the single output single input case all allocative inefficiency is scale inefficiency". Depending on the combination of these effects allocative profit efficiency, can be higher or lower than 1. The interpretation of allocative efficiency values will be undertaken in the next section where we show how allocative efficiency can be decomposed into mix and scale effects and interpret each of

[5]Note that model (8.6) is also similar to the preference model introduced by Thanassoulis and Dyson (1992) where we assume that there are no preferences. If there are any preferences for moving towards the frontier these can be incorporated in the model above as detailed in Thanassoulis and Dyson (1992).

these components.

Phase 2 - Decomposing Allocative Profit Efficiency

The allocative profit efficiency measure may reflect changes in the mix of inputs and outputs and changes in the scale of operations when moving from the technically efficient point $(\mathbf{x}', \mathbf{y}')$ to the maximum profit point $(\mathbf{x}^*, \mathbf{y}^*)$. Analysing first changes in scale, consider a situation where increasing inputs (say by the scaling factor α) generates a less than proportional increase in outputs (say scaling them by γ, $\gamma < \alpha$). This would imply we have decreasing returns to scale. This means that if only scale effects were present in moving from $(\mathbf{x}', \mathbf{y}')$ to $(\mathbf{x}^*, \mathbf{y}^*)$, then the allocative profit efficiency component would be $\frac{(\Pi_i \frac{x_i^*}{x_i'})^{1/m}}{(\Pi_r \frac{y_r^*}{y_r'})^{1/s}}$ $= \frac{(\Pi_i \frac{\alpha x_i'}{x_i'})^{1/m}}{(\Pi_r \frac{\gamma y_r'}{y_r'})^{1/s}}$, which equals $\frac{\alpha}{\gamma} \geq 1$. The scale change effect can therefore be measured through $\frac{\alpha}{\gamma}$. In order to isolate this effect we propose the use of model (8.7), where the scale change effect is minimised.

$$\min \left\{ \alpha/\gamma \mid y_r^* - s_r = \gamma y_r' \;\; r = 1, \ldots, s, \;\; x_i^* + e_i = \alpha x_i' \;\; i = 1, \ldots, m, \;\; s_r, e_i \geq 0 \right\} \quad (8.7)$$

In constructing model (8.7) we assumed that scale effects resulting of movements from the technically efficient point to the maximum profit point (which is also technically efficient) exist when we can increase (decrease) all inputs proportionally and all outputs proportionally so that the maximum profit point has higher outputs and lower inputs than the scale adjusted point: $y_r^* \geq \gamma y_r'$ and $x_i^* \leq \alpha x_i'$. This means that the scale adjusted point will lie on the free disposable hull of the maximum profit point, since at least one input and one output will be equal to that of the maximum profit point with the remaining inputs being no lower and the remaining outputs being no higher.

The concept of scale implicit in (8.7) relates to the possibility of exploiting scale economies in movements between two technically efficient points on the production frontier. The concept of scale economies is closely related to the concept of returns to scale, but none of these concepts is free of controversy in the literature as testifies the works of Gold (1981) and Tone and Sahoo (2003). In this Chaptrer we adopt the typical returns to scale concept of economic theory that defines such returns with reference to scale changes which maintain the mix of inputs and of outputs (Gold, 1981).

Model (8.7) provides 'radial' (measured through α and γ) and 'non-radial' adjustments (measured through s_r and e_i) required to move from the technical efficient point to the maximum profit point. The radial adjustments reflect scale change effects (given directly by α/γ), while the non-radial adjustments reflect changes in the mix of inputs and outputs. (Model (8.7) is analogous to that presented in Cooper et al. (1996) or Golany and Yu (1997) with the only difference being that we restrict the left hand side to be the maximum profit point[6].) Note that in order for scale effects to exist one needs to have both α and $\gamma \geq 1$ or both α and $\gamma \leq 1$. In fact having simultaneously a value of $\gamma > 1$ and $\alpha < 1$ implies a movement from the technically efficient point to a point outside the production possibilities set, which is not a viable movement since the scale adjusted point would not lie on the free disposal hull of the maximum profit point. Having simultaneously a value of $\gamma < 1$ and $\alpha > 1$ results in a movement towards technical inefficiency, that reduces profit rather than increasing it, which is not consistent with the profit maximisation objective. Therefore, if it is not possible to find in (8.7) values of α and γ that point to movements in the same direction, then one would conclude for the non-existence of scale effects in moving from a technical efficient point to a maximum profit point and therefore all allocative efficiency would be due to mix changes.

Turning now to the interpretation of the values of the two components we begin by assuming that scale effects as given by α/γ do exist. In this circumstance if α/γ is higher than 1 then it means that in moving from the technical efficient point to the maximum profit point the assessed unit would experience an average increase in inputs that is more than proportional to the average increase in outputs. Thus decreasing returns to scale are implicit in this movement. If the value α/γ is lower than 1, then increasing returns to scale are implicit in this movement.

The adjustment of the technically efficient levels of inputs and outputs by the scale factors α and γ, respectively leads the technically efficient unit to a scale adjusted point where $y_r^* \geq \gamma y_r'$ and $x_i^* \leq \alpha x_i'$. As the maximum profit point has higher outputs and lower inputs than the scale adjusted point then the mix effect (reflecting the GDF distance between these two points) cannot be higher than 1. The mix effect will therefore vary between 0 and 1, representing the average change in inputs and outputs that is required to move from a scale adjusted point to the maximum profit point. Therefore, for the situation where scale effects do exist an allocative profit efficiency, which is the product of scale

[6]Note that model (8.7) here is not being used for identifying returns to scale (RTS) or defining scale efficiency [because the maximum profit point may not be an mpss (see for details Cooper et al., 2000, pp. 127-131)] but simply to identify scale changes in movements from one point to the other. Indeed, RTS are a local characteristic requiring small γ and α factors. In our case one or more of these factors might be large depending on the two points being considered.

and mix effects, when greater than 1 would be the result of a scale effect greater than 1. Such a scale effect is not favourable for the production unit from the pure technological perspective, but may be so from a profit perspective. This issue will be further addressed in the empirical illustration shown in the next section.

When scale effects cannot be identified from the solution of model (8.7), all allocative efficiency is due in this case to changes in the mix of inputs and outputs. Therefore, an allocative efficiency measure $\left(\frac{(\Pi_i \frac{x_i^*}{x_i})^{1/m}}{(\Pi_r \frac{y_r^*}{y_r})^{1/s}}\right)$ greater than 1 means that the geometric average change in inputs is higher than the geometric average change in outputs in moving from the technical efficient point to the maximum profit point. This means that this movement in fact implies a decrease in the physical productivity of the unit being assessed as will become clear in the next section. Such movements may not be advisable from a pure technological perspective, though they are so from a profit perspective. Values of the allocative efficiency lower than 1 mean that the geometric average change in inputs is lower than the geometric average change in outputs and therefore such movements are advisable both from a profit and from a technological perspective.

8.4 Illustration of the GDF Measure and its Decomposition

For illustrating the calculation of overall profit efficiency and its decomposition, we will use the data shown in Table 8.1. This is the data used in Ali and Seiford (1993b), except for the hypothetical input/output prices, which we have added.

Table 8.1: Illustrative Data for units producing one output using two inputs

	Unit	1	2	3	4	5	6	7	8	9	10	11
Output	y	12	14	25	26	8	9	27	30	31	26.5	12
Input 1	x_1	5	16	16	17	18	23	25	27	37	42	5
Input 2	x_2	13	12	26	15	14	6	10	22	14	25	17
Price of y	p	20	22	24	25	23	19	18	21	23	22	25
Price of x_1	w_1	6	7	5	4	5.5	6.5	7.5	8	4	5.5	6
Price of x_2	w_2	9	8	7.5	6	8	5	9	10	10.5	9.5	8.5

We used model (8.2) and the GDF profit efficiency measure defined in (8.3) on the data in Table 8.1. The results are presented in Table 8.2. In this table we also show results for the numerator and denominator of the GDF measure, called input change and output change, respectively.

There are two units that are overall profit efficient (unit 7 and 9). All the other units

Table 8.2: Overall Profit Efficiency Measurement

Unit	Actual Profit	Maximum Profit	Peer Unit	Input Change	Output Change	GDF Profit Efficiency
1	93	300	7	1.96	2.25	87.16%
2	100	339	7	1.14	1.93	59.17%
3	325	454	9	1.12	1.24	89.99%
4	492	543	9	1.43	1.19	119.54%
5	-27	403.5	7	0.99	3.38	29.51%
6	-8.5	308.5	4	1.36	2.89	47.05%
7	208.5	208.5	7	1.00	1.00	100%
8	194	267	7	0.65	0.90	72.08%
9	418	418	9	1.00	1.00	100%
10	114.5	361.5	7	0.49	1.02	47.89%
11	125.5	440	7	1.72	2.25	76.22%

are overall profit inefficient since they fail to maximise profits given their input and output prices.

Most units in Table 8.2 achieve profit efficiency by increasing simultaneously inputs and outputs, the latter increment being on average more than proportional to the former. (Unit 8 is an exception, as it shows no increase but decrease in both inputs and outputs). Note that unit 4 is the only one that should on average increase outputs less than proportionally to inputs. Such a result is related with the type of returns to scale that apply at this point as will become clear later.

Decomposition of Overall Profit Efficiency

The application of the 4-step procedure detailed in the previous section to identify technically efficient targets resulted, in its first step, in a set E consisting of units {1, 3, 4, 6, 7, 8, and 9}. Qhull was then applied to find the efficient facets. These are $F_1 = \{1,6,7\}$, $F_2 = \{1,3,4\}$, $F_3 = \{1,4,7\}$, $F_4 = \{4,8,9\}$, $F_5 = \{4,7,9\}$. All the facets in this case are FDEF, and there are no NFDEF. In Figure 8.2 we represent graphically the above facets. Note that the facets {3, 8, 4} and {6, 7, 9} are not efficient, but they are identified in the graphs of Ali and Seiford (1993b) as being Pareto-efficient. Therefore, there is no full correspondence between our Figure 8.2 and the graphs in Ali and Seiford (1993b).

The results from applying model (8.6) to our illustrative example are shown in Table 8.3, where only inefficient units are considered. For example, for unit 5 only projections on F_1, F_2, and F_3 were feasible (note that in technical efficiency measurement we do not allow units to decrease outputs nor to increase inputs, meaning that some facets might not

Figure 8.2: Illustrative Example

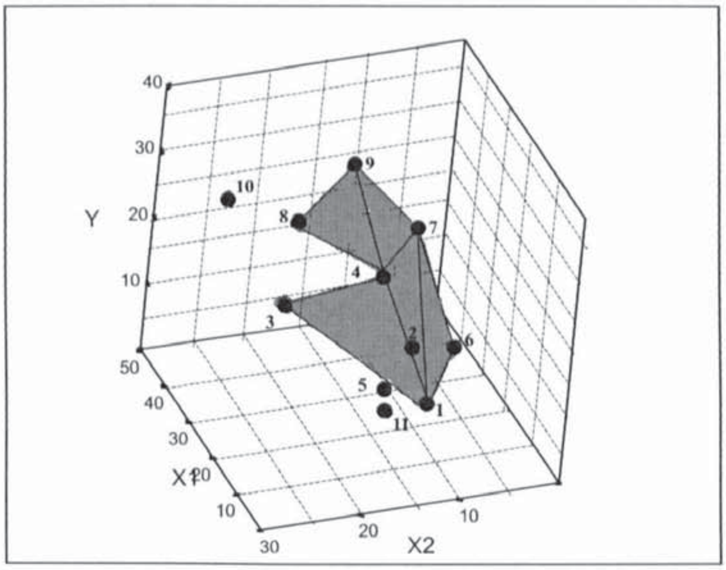


Table 8.3: Technical Efficiency Measurement Results

Unit	2	5	10	11
Eff. Facet	F_1	F_1	F_4	F_1, F_2, F_3
Observed (y, x_1, x_2)	(14, 16, 12)	(8, 18, 14)	(26.5, 42, 25)	(12, 5, 17)
Model (8.6) targets	(14, 16, 9.72)	(9.83, 18, 7.94)	(30, 27, 22)	(12, 5, 13)
GDF	90%	61.3%	66.4%	87.5%
Peer Units	$\lambda_1 = 0.41, \lambda_6 = 0.38, \lambda_7 = 0.21$	$\lambda_1 = 0.28, \lambda_6 = 0.72$	$\lambda_8 = 1$	$\lambda_1 = 1$
Additive targets	(17, 11.67, 12)	(19, 11, 14)	(26.5, 21, 12.5)	(12, 5, 13)
GDF	70.3%	32.9%	50%	87.5%

be attainable from the actual position of some units). F_1 was chosen because in this case the objective function of (8.6) was 0.6617 as opposed to a value of 1.294 when F_2 or F_3 were chosen. Note that we also show in Table 8.3 the results from the additive model when it is used for finding technically efficient targets, and present the GDF measure for both situations. Clearly our model provides targets that are closer to the inefficient units being assessed as can be confirmed by the higher GDF technical efficiency measures presented in Table 8.3 or by calculating distance measures such as L_p norms (see Portela et al., 2003).

Allocative profit efficiency results for our illustrative example are shown in Table 8.4, where we also present the type of RTS that apply at each technically efficient point (RTS were identified through the procedure of Färe et al. (1985)). Knowing the RTS characteristics of each unit we can now easily understand why unit 4 has an allocative profit efficiency greater than 100%. Unit 4 lies on a region of CRS and must move, in order to be profit

Table 8.4: Allocative Profit Efficiency Measurement Results

Unit	1	2	3	4	5	6	7	8	9	10	11
All.Eff.	87.16%	65.74%	89.99%	119.54%	48.15%	47.05%	100%	72.08%	100%	70.93%	87.16%
RTS	CRS	IRS	DRS	CRS	IRS	IRS	CRS	DRS	DRS	DRS	CRS

Table 8.5: General Efficiency Measurement Results for Some Units

Unit	Overall Profit GDF	Technical GDF	Allocative GDF	Scale effect	Mix effect
2	59.17%	90.00%	65.74%	1.56/1.93 = 0.8102	0.8114
3	89.99%	100%	89.99%	2.31/1.24 =1.865	0.4825
4	119.54%	100%	119.54%	2.18/1.19=1.825	0.655
5	29.51%	61.3%	48.15%	1.39/2.75=0.506	0.952
6	47.05%	100%	47.05%	2.5/2.89 = 0.865	0.544

efficient, to unit 9, a DRS unit. As it is well known CRS points are most productive scale size (mpss) in the sense of Banker (1984). This means that by comparing a mpss unit such as 4 with a non-mpss unit such as 9 the productivity will decrease. For the specific case of unit 4 we have partial productivity ratios at its original position of: $\frac{y}{x_1} = \frac{26}{17}$ and $\frac{y}{x_2} = \frac{26}{15}$, and at the maximum profit point (unit 9) partial productivity ratios are: $\frac{y}{x_1} = \frac{31}{37}$ and $\frac{y}{x_2} = \frac{31}{14}$. The geometric mean of these productivity ratios is 1.628183 ($\sqrt{26/17 \times 26/15}$) at the original position of unit 4, and 1.362062 ($\sqrt{31/37 \times 31/14}$) at its maximum profit position. The ratio between these two values (1.628183/1.362062) gives 119.54%, indicating that the mean productivity of unit 4 is higher than that at its maximum profit target point. Note that an allocative profit efficiency measure greater than 1 will always indicate a productivity change that is not beneficial in strict technical terms for the unit being assessed, although it is so in profit terms, because of the relative input/output prices.

Allocative profit efficiency was decomposed into its components through model (8.7). In Table 8.5 we show the results, for some units, on the various efficiency measures including the scale and the mix components estimated as explained earlier. Only three units, 2, 5 and 6, present scale effects below 1. This is due to the IRS characteristics of the technically efficient projections of these units. As the maximum profit targets for these units are all located in CRS, they can exploit the IRS potential by increasing outputs more than proportionally with inputs (i.e. $\gamma \geq \alpha \Rightarrow \alpha/\gamma \leq 1$).

To illustrate the calculation of mix effects take for example unit 5, whose technically efficient target is (y, x_1, x_2)=(9.83, 18, 7.94). If inputs are multiplied by 1.39, outputs will increase by 2.75, and thus the scale adjusted point is (27, 25, 11.034). As the maximum profit point is (27, 25, 10) the GDF between this point and the scale adjusted point is the mix effect, which equals 0.952.

8.5 Profit Efficiency when some Prices are Unknown

In the previous exposition of the GDF approach we assumed that prices were known for each factor included in the computation of maximum profit (model (8.2)). In practice, however, some prices may be unknown. In this situation, a commonly used assumption is that prices are equal for all production units (e.g. Färe et al., 1990). Being this so, the calculation of maximum profit can be equivalently undertaken through the use of price and quantity information (when prices are known), or cost and revenue information (when prices are unknown).

Under the assumption of equal prices for all units the maximum profit model [see (8.2)] turns out to be equivalent to model (8.8), where $R_r = p_r y_r$, $C_i = w_i x_i$, $R_{rj} = p_r y_{rj}$, and $C_{ij} = w_i x_{ij}$.

$$\max_{\lambda_j, R_r, C_i} \Big\{ \sum_{r=1}^{s} R_r - \sum_{i=1}^{m} C_i \mid \sum_{j=1}^{n} \lambda_j \; R_{rj} \geq R_r, \; r = 1, \ldots, s, \\ \sum_{j=1}^{n} \lambda_j \; C_{ij} \leq C_i, \; i = 1, \ldots, m, \; \sum_{j=1}^{n} \lambda_j = 1 \Big\} \quad (8.8)$$

In a study of farm's profit efficiency where price information was not available, Färe et al. (1990) used a profit linear programming model as shown in (8.8) (see also Färe et al., 1994a, Chap. 10), where cost (represented by C) and revenue (represented by R) were used instead of quantities and prices. The authors proved that this model was equivalent to model (8.2) under the assumption of equal prices for all the units. As stated in (8.8) the model needs to be solved only once, and maximum profit is equal for all production units.

Model (8.8) corresponds to the extreme case where no prices are known for any factor. It may however happen that prices are known for some factors and unknown for others. In this situation we can think of a mixed profit maximisation model of the type shown in

(8.9), where unknown prices are assumed to be equal for all production units.

$$\max_{\lambda_j, y_r, x_i, R_r, C_i} \Big\{ \sum_{r=1}^{k} p_{ro} y_r - \sum_{i=1}^{l} w_{io} x_i + \sum_{r=k+1}^{s} R_r - \sum_{i=l+1}^{m} C_i \mid \sum_{j=1}^{n} \lambda_j \, y_{rj} \ \geq \ y_r, \ r = 1, \ldots, k$$
$$\sum_{j=1}^{n} \lambda_j \, x_{ij} \ \leq \ x_i, \ i = 1, \ldots, l \, , \ \sum_{j=1}^{n} \lambda_j \, R_{rj} \ \geq \ R_r, \ r = k+1, \ldots, s$$
$$\sum_{j=1}^{n} \lambda_j \, C_{ij} \ \leq \ C_i, \ i = l+1, \ldots, m \, , \ \sum_{j=1}^{n} \lambda_j = 1 \Big\} \quad (8.9)$$

The equivalence between this model and the maximum profit model (8.2) is proved by considering for the factors on which price is unknown the variables $R_r = p_r y_r$ and $C_i = w_i x_i$, where p_r and w_i are equal across all units. These equalities make the objective function of (8.9) equal to the objesctive function of (8.2), and the constraints are also equal as $\sum_{j=1}^{n} \lambda_j \, R_{rj} \geq R_r \Leftrightarrow \sum_{j=1}^{n} \lambda_j \, p_r \, y_{rj} \geq p_r y_r \Leftrightarrow \sum_{j=1}^{n} \lambda_j \, y_{rj} \geq y_r$ and $\sum_{j=1}^{n} \lambda_j \, C_{ij} \leq C_i \Leftrightarrow \sum_{j=1}^{n} \lambda_j \, w_i \, x_{ij} \leq w_i x_i \Leftrightarrow \sum_{j=1}^{n} \lambda_j \, x_{ij} \leq x_i$.

Maximum profit targets are useful for calculating the profit efficiency measure, which is usually disentangled into its technical and allocative components. The use of a 'mixed' maximum profit model does not pose a problem here as the variables to be used in the technical and allocative efficiency models are the quantities of inputs and outputs for which price is known and the costs and revenues associated with those inputs and outputs for which price is unknown. Our technical efficiency measure in (8.6) is therefore modified to (8.10).

$$\min_{\lambda_j, \gamma_{ro}, \beta_{io}} \Big\{ \sum_{r=1}^{s} \gamma_{ro} + \sum_{i=1}^{m} \beta_{io} \mid \sum_{j \in F_k} \lambda_j y_{rj} = y_{ro} + \gamma_{ro} y_{ro}, \ r = 1, \ldots, k \, ,$$
$$\sum_{j \in F_k} \lambda_j x_{ij} = x_{io} - \beta_{io} x_{io}, i = 1, \ldots, l, \ \sum_{j \in F_k} \lambda_j R_{rj} = R_{ro} + \gamma_{ro} R_{ro}, \ r = k+1, \ldots, s,$$
$$\sum_{j \in F_k} \lambda_j C_{ij} = C_{io} - \beta_{io} C_{io}, \ i = l+1, \ldots, m, \ \sum_{j \in F_k} \lambda_j = 1, \ \lambda_j, \ \gamma_{ro}, \ \beta_{io} \ \geq \ 0 \Big\} \quad (8.10)$$

Assuming equal prices of those factors for which cost or revenue is considered, we have that $\sum_{j \in F_k} \lambda_j R_{rj} = R_{ro} + \gamma_{ro} R_{ro} \Leftrightarrow \sum_{j \in F_k} \lambda_j (p_r y_{rj}) = (p_r y_{ro}) + \gamma_{ro} (p_r y_{ro}) \Leftrightarrow p_r \sum_{j \in F_k} \lambda_j y_{rj} = p_r (y_{ro} + \gamma_{ro} y_{ro}) \Leftrightarrow \sum_{j \in F_k} \lambda_j y_{rj} = y_{ro} + \gamma_{ro} y_{ro}$, which proves the equivalence between constraints using revenues and constraints using quantities of outputs. For the input case the proof is analogous.

8.6 Constrained Profit Analysis

The maximum profit model presented previously [model (8.2)] assumes that all production factors are variable. This is a usual assumption in long run profit maximisation. In the short run it may be impossible to change some of the factors of production, and these should therefore be considered fixed (e.g. Färe et al., 1990). In the short run one may also consider other type of factors that, although changeable cannot be changed by large amounts. Take for example staff. In the short run it may be possible to add or to exclude a few staff members from a production unit but a large number of these cannot be hired or dismissed in the short run (time consuming negotiations need to take place in firing people and arrangements for a new space might take place if a large number of people is to be hired). For this type of factors it is usual to restrict some costs and/or revenues to grow no more than a certain amount in the short run.

The idea of constraining maximum profit is explored in Färe and Grosskopf (1994a) where the authors put forward cost, revenue and profit models with additional constraints. These models are called 'indirect' cost and revenue models. The cost indirect revenue function is according to Färe and Grosskopf (1994a) defined in the same way as the ordinary revenue function, but it is subject to a budget constraint that prevents input costs to grow by more than a certain amount. On the other hand, the revenue indirect cost function is defined as the ordinary cost model to which a revenue constraint is added. In a previous application to rice farms Färe et al. (1990) have used a nonparametric profit function with expenditure constraints. The use of constrained maximum profit was also put forward by Cooper et al. (2000). These authors used the additive model to calculate technical efficiency and put forward a different, but equivalent, form of model (8.2) where slacks are considered in the objective function. When these slacks (representing for each factor the difference between revenues and costs at the maximum profit point and the observed point) are allowed to vary freely the model provides unconstrained maximum profit. The authors, however, imposed bounds on these slacks "so that the resulting projections do not go far from the observed values and remain in managerially and technically allowable ranges" (Cooper et al., 2000, p. 225).

Note that the imposition of bounds on costs and revenues translates in imposing bounds on the corresponding quantities of inputs and outputs. In fact a budget constraint for two inputs could be $w_1x_1+w_2x_2 \leq \alpha(w_1x_{1o}+w_2x_{2o})$. That is, the optimal cost to be determined shall not exceed a given percentage of actual cost. This cost constraint can consider all costs aggregated (the most usual case), but it could also disaggregate costs by imposing $w_1x_1 \leq \alpha w_1x_{1o}$ and $w_2x_2 \leq \alpha w_2x_{2o}$. Obviously in this disaggregation input prices can be ignored and only quantities are left in the constraints.

Assume that constraints of the type shown in (8.11) are to be used in the maximum profit model (8.2). If α is specified in 20% it means that we wish target inputs to be no more than 20% higher than observed inputs, and no more than 20% lower than observed inputs. The same type of interpretation is valid for the factor γ associated to outputs.

$$(1-\alpha)x_{io} \leq x_i \leq (1+\alpha)x_{io} \ ; \ (1-\gamma)y_{ro} \leq y_r \leq (1+\gamma)y_{ro} \tag{8.11}$$

Constraints such as those specified in (8.11) can be directly added to the set of constraints in model (8.2).

As far as the measurement of profit technical efficiency is concerned, the imposition of constraints in (8.11) may imply projections on the interior of the original production frontier which are not Pareto-optimal. This means that the closest target model (8.6) cannot be used together with constraints on observed input and output levels, as this model forces projections on Pareto-efficient facets. As an alternative to measure technical efficiency in the presence of input and output constraints we use model (8.12).

$$\begin{aligned}\text{EFF}_o = \min \Bigg\{ \ \theta/\beta \mid \ & \sum_{j=1}^{n} \lambda_j \ y_{rj} \geq \beta y_{ro}, \ \ \sum_{j=1}^{n} \lambda_j \ x_{ij} \leq \theta x_{io}, \\ & \sum_{j=1}^{n} \lambda_j = 1, \ \sum_{j=1}^{n} \lambda_j y_{rj} \leq (1+\gamma) y_{ro}, \ \sum_{j=1}^{n} \lambda_j x_{ij} \geq (1-\alpha) x_{io}, \\ & \qquad \lambda_j, \ \geq \ 0, \ 0 \leq \theta \leq 1, \ \beta \geq 1 \Bigg\}\end{aligned} \tag{8.12}$$

The constrained model (8.12) uses an equiproportional factor associated to contracting inputs and a different equiproportional factor associated with expanding outputs. The objective function of (8.12) is therefore a special case of the GDF measure where all inputs are assumed to change by the same proportion and all outputs are assumed to change by the same proportion. The above model can result in some positive slacks, which can be accounted for by calculating the GDF *a posteriori* using the targets resulting from (8.12). Note that, as we are measuring technical efficiency, we retain in (8.12) the original assumption that inputs cannot be expanded and outputs cannot be contracted towards technical efficiency. This means that only upper bounds are required on output changes and only lower bounds are required on input changes. Note that the last constraints restricting input and output changes result in β values that are no larger than $(1+\gamma)$ and θ values that are no lower than $(1-\alpha)$.

8.7 Summary

In this Chapter we review some of the main approaches used in profit efficiency measurement and propose a novel way of calculating profit efficiency. A geometric distance function, based on geometric means of productivity ratios, is introduced as a means to calculate overall profit efficiency and its technical and allocative components. For that purpose all that is needed is the profit maximising point and a Pareto-efficient point corresponding to each unit. The former allows the calculation of overall profit efficiency and the latter allows the calculation of technical profit efficiency. Allocative profit efficiency can then be found by decomposition. The above procedure is valid whatever the means used to find the profit maximising and the Pareto-efficient points. For finding the Pareto-efficient targets we propose, however, a procedure [based on Portela et al. (2003)] that aims at finding the nearest to the unit being assessed rather than the furthest point on the Pareto-efficient frontier.

We examine some properties of the profit efficiency measure defined and also show that allocative profit efficiency can be decomposed into scale and mix effects. We propose a way of disentangling these effects, which throws light on the understanding of allocative efficiency in a profit context. We recognise, however, that more needs to be done on this issue.

The GDF approach is based on a concept of absolute maximum profit that may provide difficult-to-achieve targets in the short run. A natural extension of our procedure is therefore to calculate relative maximum profit [as proposed in Cooper et al. (2000, p. 225)] and measure overall profit efficiency in relation to this. Such an extension is also analysed in this Chapter.

Chapter 9

Productivity Change and Malmquist Indexes based on the GDF Efficiency Measure

In this Chapter we analyse in some detail the use of the GDF efficiency measure developed in the previous Chapter to the calculation of productivity change and Malmquist indexes. For this purpose we focus on this Chapter on the use of the GDF for technical efficiency measurement rather than for economic (profit) efficiency measurement. We start this Chapter by providing some insights regarding the computation of Malmquist indexes and then we show some drawbacks of existing approaches to calculate productivity change through Malmquist indexes. The approach we propose based on the GDF tries to circumvent the identified problems.

9.1 Introduction

Productivity change has been a topic of interest since the earlier developments on this matter by Caves et al. (1982) on Malmquist productivity indexes. Earlier in 1978 the work of Charnes, Cooper and Rhodes on DEA provided a straightforward way to measure efficiency through linear programming models, and since then this framework has been applied to the measurement of productivity change through Malmquist indexes. Malmquist indexes, using DEA efficiency measures calculated in relation to a constant returns to scale (CRS) technology, are argued to be equivalent to a total factor productivity (TFP) index (see e.g. Färe et al., 1994b, 1998). This is easily proved for single input/output technologies, but for multiple input/output technologies the calculated TFP Malmquist index has some problems. In this Chapter we refer in particular to problems arising from the fact that a reference technology is used relative to which technical efficiency is assessed. Indeed, the TFP Malmquist index computes productivity change between two observed points, say **a** and **b**, by finding one or more reference point relative to which the technical efficiency of **a** and **b** is assessed. Productivity change is then inferred from changes in technical efficiency. We argue in this Chapter that TFP should be measured by comparing directly the points **a** and **b** rather than using references that might not be the same for each point.

This Chapter proposes, therefore, a novel way to compute TFP using observed values only, which does not require any specifications about the technology on which points such as **a** and **b** operate. The proposed TFP measure is then decomposed into efficiency change, technological change, and a residual effect which reflects scale and allocative shifts. This decomposition obviously requires assumptions about the technology under which production units operate. In our TFP decomposition we try to account for some problems on existing methodologies such as the FGNZ approach of Färe et al. (1994b) and the RD approach of Ray and Desli (1997) (see Chapter 2 section 2.3.6). Both approaches calculate the TFP Malmquist productivity index in the same way (through radial efficiency scores calculated in relation to a CRS technology), but they decompose it differently. In the FGNZ approach the technological change component is calculated with reference to a CRS frontier, while in the RD approach it is calculated with reference to a VRS (variable returns to scale) frontier. Therefore, the FGNZ approach has the advantage of measuring changes in "maximal average product" (Färe et al., 1997b), but it has the disadvantage of not accounting for changes in the VRS technology. This might be a serious drawback if there are strong reasons to believe that the true technology is indeed VRS (see e.g. Balk, 2001). The RD approach tries to solve the problem of the FGNZ approach by specifying a technological change component that is defined in relation to a VRS technology. This, however, may result in some computational problems because some DEA models might be infeasible when assessments involve cross-period data (Bjurek, 1996).

Both the FGNZ and the RD approaches are based on radial efficiency measures that are oriented either towards input contraction or output expansion. This provides different results concerning some components of productivity change depending on the model orientation. The orientation of DEA models is in some cases a given since some inputs and outputs are not under the control of production units. However, in many cases (for example in the assessment of bank branches as we undertake in our empirical application in Chapter 10) at least some inputs and some outputs are under the control of production units and in such cases non-oriented models might be used instead. The use of non-oriented efficiency measures solves the problem of sensitivity of the solution to the model's orientation, while at the same time solving the computational problems inherent to the RD approach. Examples of non-oriented efficiency measures that have been used in this context are the directional distance function used by Chambers et al. (1996b) and Chung et al. (1997), and the hyperbolic efficiency measure used by Zofio and Lovell (2001).

Another problem of the FGNZ and RD approaches to calculate Malmquist indexes is that they rely on radial measures that do not account for slacks. If slacks are important sources of inefficiency, then the resulting Malmquist indexes may be based on biased measures of efficiency that do not fully reflect the distance between observed values and targets. Some authors have addressed this problem and propose to solve it through the use of non-radial efficiency measures (note that non-radial efficiency measures are not necessarily non-oriented, though the reverse is true). For example, Grifell-Tatjé et al. (1998) developed what they called a quasi-Malmquist productivity index that tries to overcome this problem (see also Førsund, 1998, who criticise this paper), and Thrall (2000) developed an efficiency measure (based on a weighted additive model) that can be used in the computation of Malmquist type indexes.

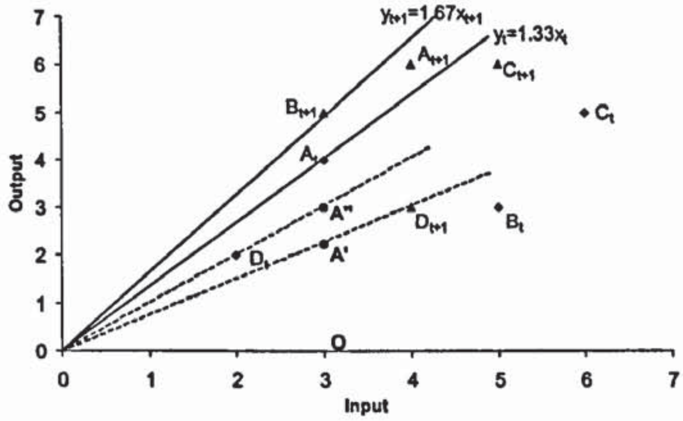
The measure of efficiency that we use in this Chapter to decompose TFP is the GDF measure detailed in the previous Chapter. This is a non-oriented efficiency measure and is able to account for all the sources of inefficiency, therefore avoiding the above mentioned problems.

9.2 Problems with Traditional ways of Calculating TFP

The FGNZ and RD approaches use radial efficiency measures calculated in relation to CRS frontiers to calculate Malmquist TFP indexes. Careful reflection shows, however, that this is just a means to an end since productivity change, reflected in TFP measures, concerns observed values only and does not require the use of efficiency measures.

Consider the single input/output case where a measure of productivity change from period t to $t+1$ is given by the ratio $\frac{P_{t+1}}{P_t}$, where $P_t = \frac{y_t}{x_t}$ in each time period t. This productivity change measure is put forward by most authors that analyse productivity change and is free of controversy. Graphically $\frac{P_{t+1}}{P_t}$ corresponds to the distance between the rays that pass through a given observation in period t and $t+1$ (see Figure 9.1). The highest of these rays in each time period is the CRS frontier of that period (associated with highest productivity). Obviously the distance between the rays that pass through, for

Figure 9.1: Single Input/Output Example



example, points D_{t+1} and D_t in Figure 9.1 can be alternatively calculated with reference to another ray. Taking this reference as being the CRS frontier of period t defined by unit A_t, we have that the distance between the rays that pass through D_{t+1} and A_t divided by the distance between the rays that pass through D_t and A_t is equal to the distance between the rays that pass through D_{t+1} and D_t (in the graph this is the same as to say $\frac{OA'}{OA''} = \frac{OA'/OA_t}{OA''/OA_t}$). Based on this fact, existing approaches propose the use of distance functions defined in relation to CRS technologies to calculate productivity change indexes for the general case of multiple inputs/multiple outputs. Such distance functions are of the type introduced by Farrell (1957), which are usually operationalised through DEA (note that in Figure 9.1 the ratios OA'/OA_t and OA''/OA_t are the Farrell output efficiency measures of units A' and A'', respectively).

Consider a measure $\gamma_{j_t}^t$ indicating the radial efficiency of unit j as observed in period t and assessed in relation to the technology of period t (superscript). A Malmquist productivity index, M_j^t, is usually computed as $\frac{\gamma_{j_{t+1}}^t}{\gamma_{j_t}^t}$, when the reference is the t frontier. Obviously the reference technology could also have been $t+1$, which would result in $M_j^{t+1} = \frac{\gamma_{j_{t+1}}^{t+1}}{\gamma_{j_t}^{t+1}}$. The values of these two Malmquist indexes may differ and, as such, Färe et al. (1994b) consider the geometric mean of both as the Malmquist total factor productivity index as shown in (9.1).

$$M_j = \left(\frac{\gamma_{j_{t+1}}^t}{\gamma_{j_t}^t} \times \frac{\gamma_{j_{t+1}}^{t+1}}{\gamma_{j_t}^{t+1}}\right)^{(1/2)} \tag{9.1}$$

Table 9.1: Illustrative Example

	Period t			Period $t+1$			Growth	
Unit	y	x_1	x_2	y	x_1	x_2	$\Delta(y/x_1)$	$\Delta(y/x_2)$
Unit 1	12	5	13	22	8	14	1.146	1.7024
Unit 2	14	16	12	12	12	11	1.143	0.9351
Unit 3	26	16	26	26	8	25	2	1.0816
Unit 4	26	17	15	20	15	14	0.872	0.8242
Unit 5	8	12	14	8	6	10	2	1.4

Note that productivity change, as shown in Figure 9.1, is not dependent on efficiency or functional form of the efficient frontier as defined in DEA. The use of distance functions is just a means to operationalise the concept for the multiple input/output case. This approach relies, however, on efficiency being calculated in relation to a unique referent line or plane. This necessarily happens in the single input/output case as the ray presenting maximum productivity in each time period is unique. If the referent hyperplane is not the same for observations in t and $t+1$, then the Malmquist index as defined in (9.1) is just an approximation for true productivity change and not a real measure of productivity change. In the multiple input/output case, CRS technologies are defined by a cone that has multiple facets, and projections on this cone may happen on any of its facets. This means that the referent hyperplane, or facet, is not necessarily the same for every two observations in t and $t+1$ between which productivity change is to be measured.

To illustrate this consider the example in Table 9.1, where 5 units producing one output (y) from 2 inputs (x_1 and x_2) are considered. In Table 9.1 we also show the growth in partial productivity between periods t and $t+1$. That is, calculating the partial productivity of the output in relation to input 1, y/x_1, and the partial productivity of the output in relation to input 2, y/x_2, for each time period, the ratio $\Delta\frac{y}{x_i} = \frac{y_{t+1}/x_{i_{t+1}}}{y_t/x_{i_t}}$ shows partial productivity growth of output in relation to each input i. Inspecting these ratios in Table 9.1, it is clear that units 1, 3 and 5 increased their productivity from t to $t+1$, while the productivity of unit 4 decreased in the same period. Note also, that unit 5 shows the highest productivity increase from t to $t+1$ since the partial productivity growth ratios are together the highest that can be found. If we now apply (9.1) to calculate productivity change the values are as shown in Table 9.2, where M_j is the geometric mean of M_j^t and M_j^{t+1}. These results show some contradiction to what was expected from the partial productivity ratios, especially because unit 5 does not have the highest Malmquist index as one would expect. At the same time, while it is clear that unit 4 exhibited a productivity decrease (and the Malmquist index

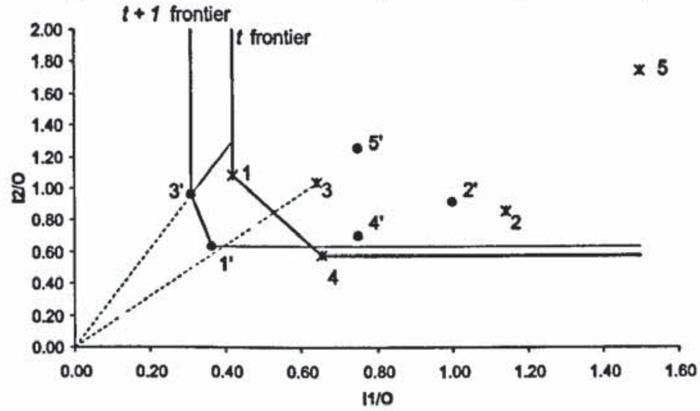
Table 9.2: Malmquist Results for Illustrative Example

Unit	γ^t_{jt}	γ^{t+1}_{jt+1}	γ^{t+1}_{jt}	γ^t_{jt+1}	M^t_j	M^{t+1}_j	M_j
Unit 1	1	1	0.7846	1.3965	1.397	1.275	1.3342
Unit 2	0.6731	0.6942	0.7424	0.6465	0.961	0.9351	0.9477
Unit 3	0.8527	1	0.6364	1.3542	1.5882	1.5714	1.5798
Unit 4	1	0.9091	1.103	0.8573	0.857	0.8242	0.8406
Unit 5	0.3984	0.5091	0.3636	0.692	1.74	1.4	1.5594

correctly identifies this decrease), it is not clear that unit 2 also had a productivity decline. In fact a guess on the productivity change of this unit would more likely be a productivity increase, because the growth on $\Delta(y/x_1)$ is higher than the decline in $\Delta(y/x_2)$.

The reasons for the above behaviour of the Malmquist TFP index can be better explained through Figure 9.2, where observations in $t+1$ are represented by dots and observations in t are represented by crosses. It is clear in Figure 9.2 that the hyperplane against

Figure 9.2: Illustrative 2 Inputs 1 Output Example



which efficiency is measured is not necessarily the same for an observation in t and $t+1$. For example, when unit 3 is evaluated in relation to the period t frontier it happens that unit 3 as observed in t is projected on the hyperplane defined by units 1 and 4, but unit 3 observed in $t+1$ is projected on the hyperplane of the t frontier defined only by unit 1 (facet where free disposability applies, or weakly efficient facet). The same reasoning is valid for other units in Figure 9.2.

Note also that the Malmquist index of units 2, 4, and 5 as evaluated in relation to the $t+1$ frontier (M^{t+1}_j) in Table 9.2 is exactly equal to the partial productivity change of input 2 ($\Delta(y/x_2)$) in Table 9.1. This means that when productivity change for these units is evaluated in relation to the $t+1$ frontier one of the inputs (in this case input 1) is completely neglected in the analysis. Such a result is due to projections on the 'flat' part of the frontier of $t+1$ in Figure 9.2, that satisfy free disposability of input 1. This fact strengthens what was previously said about the importance of using efficiency measures that account for all sources of inefficiency. The radial measures shown in Table 9.2 do not account for the slacks on input 1 that exist for units 2, 4, and 5 when these are projected on the $t+1$ frontier and, therefore, the Malmquist indexes based on these measures cannot account for productivity changes in input 1 but only in input 2.

In summary, the non-existence of a single referent hyperplane against which efficiency is measured for the same unit in different time periods, causes biased results on Malmquist TFP indexes that are based on such measures. In the next section we propose a GDF based approach that attempts to solve some of the problems seen hitherto.

9.3 Malmquist Type Indexes Based on the GDF

The GDF measure defined in Chapter 8 has a double role in this Chapter. On the one hand it is used to calculate efficiency measures that are non-oriented and account for all sources of inefficiency, and on the other hand it is used to calculate a TFP index based on observed values only. This TFP is then decomposed into three components, namely efficiency change (EFCH), technological change (THCH), and a residual effect (RES) in the way shown in (9.2).

$$TFP = EFCH \times THCH \times RES \tag{9.2}$$

The way each of the above terms is computed through the GDF is presented next.

9.3.1 Calculating TFP

Since productivity change depends only on observed values, we do not need efficiency measures to calculate it as long as there is a meaningful way of aggregating input and output changes. The GDF provides such a meaningful way.

Although the GDF has been originally proposed in the previous Chapter as a way to measure efficiency, it can be adapted to the present context to calculate productivity change. This is shown in (9.3), where the input/output levels considered are not observed versus targets as in (8.3) in Chapter 8 but observed in t versus observed in $t+1$.

$$TFP - GDF(x_t, y_t, x_{t+1}, y_{t+1}) = \frac{(\Pi_r \frac{y_{r_{t+1}}}{y_{r_t}})^{1/s}}{(\Pi_i \frac{x_{i_{t+1}}}{x_{i_t}})^{1/m}}. \tag{9.3}$$

For the single input/output case is is easy to see that (9.3) corresponds to a TFP index. In the multiple input/output case the GDF is a ratio between a geometric mean of output growth and a geometric mean of input growth, which is in fact a TFP index[1].

If we apply (9.3) to the illustrative example shown in section 9.2, the total factor productivity values are those shown in Table 9.3. Productivity growth is identified for

Table 9.3: TFP Results for Illustrative Example based on the GDF

	Unit 1	Unit 2	Unit 3	Unit 4	Unit 5
TFP-GDF	1.3967	1.034	1.4999	0.84765	1.6733

all units except unit 4 as expected. Note also that unit 5 is now the unit that offers the highest productivity growth, exactly as one would expect from the knowledge of the partial productivity ratios calculated previously. The GDF seems therefore a good alternative to calculate total factor productivity change, having the advantage of relying only on observed values and making, therefore, no assumptions about the technology.

9.3.2 Efficiency Change and Technological Change Components

The GDF in (9.3) used to calculate TFP is not an efficiency measure as it does not account for distances between observed and target levels but between two points observed in different time periods. In this sense the calculation of TFP does not require any assumptions regarding the technological specification of the frontier. However, when the GDF is used to calculate the efficiency change and technological change components of TFP such assumptions about the technology are required.

Consider $GDF(x_t, y_t) = \frac{(\Pi_i \theta_i)^{1/m}}{(\Pi_r \beta_r)^{1/s}}$, where $\theta_i = x_i^*/x_i$ and $\beta_r = y_r^*/y_r$ (see also (8.3) in Chapter 8). A measure $GDF^t(x_t, y_t)$ represents the efficiency measure of the input output vector $(\mathbf{x}, \mathbf{y})$ as observed in period t and projected against technology of period t (superscript). A Malmquist type index based on the GDF is given by (9.4).

$$MGDF = \left(\frac{GDF^t(y_{t+1}, x_{t+1})}{GDF^t(y_t, x_t)} \times \frac{GDF^{t+1}(y_{t+1}, x_{t+1})}{GDF^{t+1}(y_t, x_t)}\right)^{\frac{1}{2}} \tag{9.4}$$

Similarly to other approaches, this index can be decomposed in efficiency change (EFCH)

[1]See e.g. Diewert and Nakamura (2003, p. 148) who defined total factor productivity as the ratio between a measure of output growth and input growth in the multiple input/output case, however defining differently each aggregate measure of growth.

and technological change (THCH) as shown in (9.5).

$$MGDF = \frac{GDF^{t+1}(y_{t+1}, x_{t+1})}{GDF^{t}(y_t, x_t)} \times \left[\frac{GDF^{t}(y_{t+1}, x_{t+1})}{GDF^{t+1}(y_{t+1}, x_{t+1})} \times \frac{GDF^{t}(y_t, x_t)}{GDF^{t+1}(y_t, x_t)}\right]^{1/2} \tag{9.5}$$

That is,

$$MGDF = EFCH \times THCH \tag{9.6}$$

As the GDF is a general measure, the above decomposition is also general and encompasses as special cases other decomposition approaches in the literature. For example when input or output oriented CRS models are used to calculate target points implicit in the GDF, the above reduces to the FGNZ approach. Note, however, that MGDF is not necessarily equal to TFP as it is usually assumed in the literature. We consider that MGDF is simply the product of efficiency change and technological change. TFP as calculated in section 9.3.1 includes these components but may also include other components as will become clearer later.

Assume now a single input/output case, where technical efficient projections of each observation are identified by the superscript $*t$ if the projection lies on the frontier of period t and by $*t+1$ if the projection lies on the frontier of period $t+1$. In this case, each of the MGDF components assumes the form shown in (9.7).

$$EFCH = \frac{\frac{x_{t+1}^{*t+1}/x_{t+1}}{y_{t+1}^{*t+1}/y_{t+1}}}{\frac{x_t^{*t}/x_t}{y_t^{*t}/y_t}}, \quad THCH = \left(\frac{\frac{y_{t+1}^{*t+1}}{x_{t+1}^{*t+1}}}{\frac{y_{t+1}^{*t}}{x_{t+1}^{*t}}} \times \frac{\frac{y_t^{*t+1}}{x_t^{*t+1}}}{\frac{y_t^{*t}}{x_t^{*t}}}\right)^{\frac{1}{2}} \tag{9.7}$$

The EFCH component in (9.5) and (9.7) is interpreted in the usual way, i.e. when it is higher than one the efficiency of observation in $t+1$ evaluated in relation to the $t+1$ frontier (measured for the single input output case as $\frac{x_{t+1}^{*t+1}/x_{t+1}}{y_{t+1}^{*t+1}/y_{t+1}}$) is higher than the efficiency of observation in t evaluated in relation to the t frontier, and therefore there was an efficiency increase from t to $t+1$ (when EFCH is lower than one there was an efficiency decrease in moving from period t to period $t+1$). In the same way a THCH component higher than one means technological progress and a THCH component lower than one means technological regress from t to $t+1$. Note that the technological change component may be re-organised so that we have a product of input change (ICH) and output change (OCH). That is $THCH = (\frac{x_{t+1}^{*t}}{x_{t+1}^{*t+1}} \times \frac{x_t^{*t}}{x_t^{*t+1}})^{\frac{1}{2}} \times (\frac{y_{t+1}^{*t+1}}{y_{t+1}^{*t}} \times \frac{y_t^{*t+1}}{y_t^{*t}})^{\frac{1}{2}}$. An input change factor greater than 1 means that the frontier at t has higher inputs than the frontier at $t+1$. That is, there was an improvement (decrease) in inputs in moving from t to $t+1$. If the output change is higher than 1, it means that outputs in $t+1$ are on average higher than outputs in t, which also means an improvement in outputs in moving from t to $t+1$. So progress is assured when both input and output changes are greater than 1. Obviously one may have movements in different directions and in this case the resulting technological progress or regress will depend on which factor dominates the other. Note that the input and output change components of technological change are closely related to the input and output scale bias defined in Färe et al. (1997a) (see also Färe et al., 1998, 2001). In Appendix C we relate the input and output bias components of Färe et al. (1997a) with those identified by our approach.

In the multiple input/output case the above technological and efficiency change components are calculated as shown in (9.8), where again technological change is the product of input change and output change.

$$EFCH = \frac{\dfrac{\left(\Pi_i \theta_{i_{t+1}}^{t+1}\right)^{1/m}}{\left(\Pi_r \beta_{r_{t+1}}^{t+1}\right)^{1/s}}}{\dfrac{\left(\Pi_i \theta_{i_t}^{t}\right)^{1/m}}{\left(\Pi_r \beta_{r_t}^{t}\right)^{1/s}}}, \quad THCH = \left(\left(\Pi_i \frac{x_{i_{t+1}}^{*t}}{x_{i_{t+1}}^{*t+1}} \times \Pi_i \frac{x_{i_t}^{*t}}{x_{i_t}^{*t+1}}\right)^{\frac{1}{m}}\right)^{\frac{1}{2}} \times \left(\left(\Pi_r \frac{y_{r_{t+1}}^{*t+1}}{y_{r_{t+1}}^{*t}} \times \Pi_r \frac{y_{r_t}^{*t+1}}{y_{r_t}^{*t}}\right)^{\frac{1}{s}}\right)^{\frac{1}{2}} \tag{9.8}$$

The EFCH and THCH components in (9.8) include those existing in the literature, though being more general because they can handle situations where non-oriented models are used to calculate target levels. If both inputs and outputs change towards the technical efficient frontier, then the ratios considered in (9.8) account simultaneously for these changes. These ratios can be calculated both when targets lie on a CRS frontier or on a VRS frontier. We will use, however, only the latter technological specification for reasons that will become clearer in the next section.

9.3.3 Residual Effect

The MGDF in (9.4) can alternatively be decomposed as shown in (9.9), where it equals the product of a TFP index as calculated through the GDF (see section 9.3.1) and a residual component that is scale related.

$$MGDF = \frac{\left(\Pi_r \frac{y_{r_{t+1}}}{y_{r_t}}\right)^{1/s}}{\left(\Pi_i \frac{x_{i_{t+1}}}{x_{i_t}}\right)^{1/m}} \times \left(\frac{\left(\Pi_r \frac{y^{*t}_{r_t}}{y^{*t}_{r_{t+1}}}\right)^{1/s}}{\left(\Pi_i \frac{x^{*t}_{i_t}}{x^{*t}_{i_{t+1}}}\right)^{1/m}} \times \frac{\left(\Pi_r \frac{y^{*t+1}_{r_t}}{y^{*t+1}_{r_{t+1}}}\right)^{1/s}}{\left(\Pi_i \frac{x^{*t+1}_{i_t}}{x^{*t+1}_{i_{t+1}}}\right)^{1/m}} \right)^{\frac{1}{2}} \tag{9.9}$$

Note that it is TFP that one wants to decompose, and therefore the above is better expressed as (9.10).

$$TFP = MGDF \times \left(\frac{\left(\Pi_i \frac{x^{*t}_{i_t}}{x^{*t}_{i_{t+1}}}\right)^{1/m}}{\left(\Pi_r \frac{y^{*t}_{r_t}}{y^{*t}_{r_{t+1}}}\right)^{1/s}} \times \frac{\left(\Pi_i \frac{x^{*t+1}_{i_t}}{x^{*t+1}_{i_{t+1}}}\right)^{1/m}}{\left(\Pi_r \frac{y^{*t+1}_{r_t}}{y^{*t+1}_{r_{t+1}}}\right)^{1/s}} \right)^{\frac{1}{2}} \tag{9.10}$$

To see that the square root in (9.10) is scale related, consider for example the single input/output case, where the above (9.10) reduces to (9.11).

$$TFP = MGDF \times \left(\frac{\frac{x^{*t}_t}{x^{*t}_{t+1}}}{\frac{y^{*t}_t}{y^{*t}_{t+1}}} \times \frac{\frac{x^{*t+1}_t}{x^{*t+1}_{t+1}}}{\frac{y^{*t+1}_t}{y^{*t+1}_{t+1}}} \right)^{\frac{1}{2}} \tag{9.11}$$

The second term of this decomposition compares changes between output and input targets along the t and the $t+1$ frontier. As it is arbitrary to measure these changes on the t or on the $t+1$ frontier the geometric mean between both is taken in (9.11). As all the points considered in the square root in (9.11) are efficient points, the movements between these points (on each frontier) can only reflect the exploitation of scale economies or changes in the mix of operations.

The TFP as calculated through the GDF approach decomposes, therefore, in $MGDF$ (which includes a technological change and efficiency change components) and in a residual component (RES) that is scale related. Note that if all input and output targets are calculated in relation to a CRS technology, then in (9.11) we have $TFP = MGDF$ as the residual component would equal 1[2]. On the other hand, if target points are calculated in relation to a VRS technology, then the above decomposition in (9.11) is equivalent to the RD approach, where the residual component in (9.11) is equal to the RD scale

[2]Assume for example an input oriented CRS model, where target outputs correspond to observed outputs whatever the frontier where the observed unit is projected. This means that $y^{*t}_t = y^{*t+1}_t = y_t$, and $y^{*t}_{t+1} = y^{*t+1}_{t+1} = y_{t+1}$. At the same time target inputs are given by the function of the ray that passes through the origin and point (x_t, y_t). Let this function be given by $y_t = a_t x_t$ for period t and $y_{t+1} = a_{t+1} x_{t+1}$ for period $t+1$. Replacing this in the second term of (9.11) we have $\left(\frac{\frac{y_t/a_t}{y_{t+1}/a_t}}{\frac{y_t}{y_{t+1}}} \times \frac{\frac{y_t/a_{t+1}}{y_{t+1}/a_{t+1}}}{\frac{y_t}{y_{t+1}}} \right)^{\frac{1}{2}}$, which equals 1.

effect. In Appendix C this equivalence is proved for the single input/output case. For the multiple input/output case, the RD and the GDF approaches yield different TFP and RES components, but can yield the same efficiency change and technological change components when the same efficiency models are used under both approaches (to calculate efficiency scores in the RD model and target levels in the GDF model).

Interpreting the RD scale change factor is not easy as testifies Lovell (2001) and Ray (2001), since it is not a straightforward ratio of scale efficiency in two different periods (as happens in the FGNZ approach). However, it is not clear that the scale related component of productivity change should reflect changes in scale efficiency. For example, Lovell (2001) points out that the scale component of productivity change should reflect the influence of scale economies on productivity change rather than changes in scale efficiency. The author further points out that this contribution of scale economies to productivity change is provided by the scale component of the RD approach, whereas the contribution of the scale efficiency change of the FGNZ approach to explain scale economies is unclear. Being our residual component related with the RD approach its interpretation in terms of scale is not easy, especially in technologies using multiple inputs/outputs.

The residual effect reflects all the changes in the TFP, as calculated based on observed values only, that cannot be accounted for by efficiency change and technological change. Such residual effects are necessarily scale related. Note that in the single input/output case the square root in (9.11) can be alternatively written as: $\left(\frac{\frac{y_{t+1}^{*t}}{x_{t+1}^{*t}}}{\frac{y_t^{*t}}{x_t^{*t}}} \times \frac{\frac{y_{t+1}^{*t+1}}{x_{t+1}^{*t+1}}}{\frac{y_t^{*t+1}}{x_t^{*t+1}}}\right)^{\frac{1}{2}}$. Considering the production frontier at period t and assuming that $x_{t+1}^{*t} > x_t^{*t}$, we have that if $[y_{t+1}^{*t}/x_{t+1}^{*t}]/[y_t^{*t}/x_t^{*t}]$ is greater than 1 then the production function exhibits increasing returns to scale, if this ratio is equal to 1 we have constant returns to scale, and if it is lower than 1 we have decreasing returns (see Diewert and Nakamura, 2003, who also put forward this interpretation). Each term of the above geometric mean can therefore be interpreted as containing information on the returns to scale properties of the production frontier.

Though attractive, the above interpretation may have some problems. Note, for example, that if $x_{t+1}^{*t} < x_t^{*t}$, then a ratio of two output-input coefficients lower than 1 would indicate increasing returns while a value higher than 1 would indicate decreasing returns. Therefore the interpretation of values higher or lower than 1 is conditional to the relationship between input levels at the two points being compared. Another difficulty relates with the fact that the RES component is an aggregate measure of returns to scale on both the

t and $t+1$ frontiers. While a component on each frontier can be interpreted in the way suggested by Diewert and Nakamura (2003) a geometric mean of RTS on the t and $t+1$ frontier seems to be lacking an easy interpretation. For the multiple/input output case difficulties are even higher because movements along each production frontier may reflect, apart from scale effects, also mix effects. This means that in this case the interpretation of this factor becomes even more complicated. It is not our aim in this Chapter to deepen the analysis on the scale change component of the GDF measure. We interpret this factor simply as a residual effect that accounts for differences between TFP and a Malmquist index calculated in relation to a VRS technology. The issue of interpreting scale effects in Malmquist indexes deserves, however, further analysis especially when other than the FGNZ approach is used to calculate such effects.

9.4 Base Period Approach

The above decompositions of TFP into its technical change, efficiency change, and scale change components assumed changes from period t to period $t+1$. One may, however, calculate all these components in relation to a base period in accordance with Berg et al. (1992). Consider a base period frontier b. Adapting the procedure of Berg et al. (1992) to the GDF case, the Malmquist base period index is defined as shown in (9.12).

$$\begin{aligned} MGDFB^b(t,t+1) &= \frac{GDF^b(y_{t+1},x_{t+1})}{GDF^b(y_t,x_t)} = \\ &= \frac{GDF^{t+1}(y_{t+1},x_{t+1})}{GDF^t(y_t,x_t)} \times \left(\frac{GDF^b(y_{t+1},x_{t+1})}{GDF^{t+1}(y_{t+1},x_{t+1})} \times \frac{GDF^t(y_t,x_t)}{GDF^b(y_t,x_t)}\right) \end{aligned} \tag{9.12}$$

Where the first term corresponds to efficiency change and the second term (within brackets) corresponds to technological change. The above Malmquist index satisfies the circularity property meaning that $MGDFB^b(1,3) = MGDFB^b(1,2) \times MGDFB^b(2,3)$. This can be easily proved applying (9.12). Given this property we can simplify the calculation of Malmquist indexes by calculating for each time period efficiency change and technological change in relation to the base period. That is, the above Malmquist index can be simplified to (9.13) where the first term represents efficiency change and the second term represents technological change.

$$MGDFB^b(b,t) = \frac{GDF^b(y_t,x_t)}{GDF^b(y_b,x_b)} = \frac{GDF^t(y_t,x_t)}{GDF^b(y_b,x_b)} \times \frac{GDF^b(y_t,x_t)}{GDF^t(y_t,x_t)} \tag{9.13}$$

Applying the circularity property to the indexes in (9.13) we can obtain efficiency change and technological change from any period to any period.

In our decomposition of TFP when the base period approach is used to calculate its various components these assume the following from:

$$TFP = \frac{(\Pi_r \frac{y_{r_t}}{y_{r_b}})^{1/s}}{(\Pi_i \frac{x_{i_t}}{x_{i_b}})^{1/m}} \qquad EFCH = \frac{\dfrac{\left(\Pi_i \theta^t_{i_t}\right)^{1/m}}{\left(\Pi_r \beta^t_{r_t}\right)^{1/s}}}{\dfrac{\left(\Pi_i \theta^b_{i_b}\right)^{1/m}}{\left(\Pi_r \beta^b_{r_b}\right)^{1/s}}} \tag{9.14}$$

$$THCH = \frac{\dfrac{\left(\Pi_i \theta^b_{i_t}\right)^{1/m}}{\left(\Pi_r \beta^b_{r_t}\right)^{1/s}}}{\dfrac{\left(\Pi_i \theta^t_{i_t}\right)^{1/m}}{\left(\Pi_r \beta^t_{r_t}\right)^{1/s}}} \qquad RES = \frac{(\Pi_i \frac{x^{*b}_{i_b}}{x^{*b}_{i_t}})^{1/m}}{(\Pi_r \frac{y^{*b}_{r_b}}{y^{*b}_{r_t}})^{1/s}}$$

Replacing in (9.14) the factors θ and β by the respective ratios of target by observed values, and multiplying all the EFCH, THCH, and RES components the value of TFP is obtained. Using the above formulae we can calculate for each month t the TFP change in relation to the base period b. To obtain TFP change between any t and $t+1$ time periods all is needed is to apply the circularity property.

9.5 Calculating Technical Efficiency

The computation of the above components of TFP (for the base period and for the 'moving' period approaches) requires the calculation of efficiency measures for production units that are assessed in respect to their own period frontier but also in relation to frontiers of other periods. In this sense the DEA models to be used for measuring efficiency and technological change over time should be appropriate for providing super-efficiency scores for units lying above a different time period frontier. The procedure put forward in Chapter 8 for calculating the closest technical efficient targets is not practical when the purpose is to measure technical efficiency in a number of periods in respect to a number of frontiers. A simpler procedure, (used in our empirical application) is to use a simplified GDF measure that is incorporated on the objective function of a programming model as shown in (9.15).

$$\mathrm{EFF}_o = \min\left\{\frac{\theta}{\beta} \mid \sum_{j\in E}\lambda_j\, y_{rj} \geq \beta y_{ro}\ \ (a),\ \sum_{j\in E}\lambda_j\, x_{ij} \leq \theta x_{io}\ \ (b),\ \sum_{j\in E}\lambda_j = 1\ \ (c),\right.$$
$$\left.\lambda_j,\ \geq\ 0\ \ (d),\ \ [0\leq\theta\leq 1,\ \text{and}\ \beta\geq 1]\ \ (e)\ \text{or}\ [\theta\geq 1,\ \text{and}\ 0\leq\beta\leq 1]\ \ (f)\right\} \quad (9.15)$$

We use in (9.15) an equiproportional factor associated with expanding outputs, and a different equiproportional factor associated with contracting inputs. Though the objective function of (9.15) provides an efficiency score, we do not use it as the final efficiency measure. Instead Pareto-efficient targets resulting from model (9.15) are used for calculating the GDF technical efficiency measure for each unit. We assure that Pareto-efficient targets result from the linear combination of the λs in (9.15) by restricting the reference set to Pareto-efficient units (units in the set E).

The last set of constraints (e and f) in (9.15) assures the right direction (expansion or contraction) to be followed by inputs and outputs. When units are assessed in relation to a frontier containing observations of the same time period, only constraints (e) are activated. When the frontier relates to a different time period then two things may happen for an observation: either it lies below the frontier (and then constraints (e) are activated) or it lies above the frontier (and constraints (f) are activated). This is easily done using GAMS. Note that when a unit is above the frontier and the first set of constraints is active then the model will be unfeasible. So in GAMS the above model was simply programmed by analysing infeasibility and relax the first set of constraints in case it happened.

9.6 Summary

This Chapter draws attention to some limitations of current approaches to calculate Malmquist indexes, and attempts to solve them through the use of a geometric distance function (GDF) approach. The GDF is used here with two purposes. (i) To calculate a total factor productivity measure based on observed values only, and (ii) to calculate measures of technical efficiency that are non-oriented and account for all the sources of inefficiency. The latter use of the GDF solves the problem of infeasibility of some DEA models when VRS technologies are used, and resolves the ambiguity resulting from the use of oriented models that yield conflicting information depending on efficiency measures being input or output oriented. The former use of the GDF to calculate TFP is consistent with the single input/output case, where it is widely accepted that a ratio of productivity at two different points in time reflects productivity change. Such ratios are based on observed values only, and do not require any assumptions regarding the form of the production frontier.

TFP is decomposed in our approach into three components: efficiency change, technological change, and a residual component. This decomposition throws some light on traditional ways to decompose Malmquist indexes, and shows that more needs to be done regarding the interpretation of scale change components existing in the literature.

Chapter 10

Empirical Analysis

This Chapter applies some of the theoretical concepts outlined in previous Chapters to a sample of bank branches of a Portuguese financial group. We have identified the major dimensions of efficiency to be analysed according to the managers' views regarding the objectives of bank branches. Inside each of these dimensions managers are interested in ranking bank branches, in setting achievable targets for them, and in defining benchmark branches whose good practice can be emulated by other branches. In addition, a comparison of the performance of bank branches in each of the three dimensions identified is also of interest as managers expect that some branches perform well on one dimension (say operational efficiency), while having a poorer performance on other dimensions (say profit efficiency). This Chapter shows detailed results on each performance dimension, while the next Chapter draws together the efficiency results from the three different dimensions.

10.1 Introduction

The bank under analysis requires complete anonymity, and therefore we shall refrain from presenting any historical or background figures concerning the bank. At the same time the names of all branches analysed are hidden behind numerical codes. The sample of bank branches analysed here consisted initially of 60 branches all operating in the northern region of Portugal. Data were provided on a monthly basis starting in February 2001, though some variables (mainly relating with the transactional efficiency assessment) are available only from January 2002 onwards. Some of the initial branches in the sample were meanwhile closed and the sample progressively reduced. This means that in each month the number of branches in the sample is not exactly the same due to some closures.

Each bank branch was assessed in each month for the three efficiency dimensions that we put forth in Chapter 5, namely transactional efficiency, operational efficiency, and profit efficiency. An analysis of efficiency change and technological change over time was also performed for each performance dimension. For this purpose we used base period Malmquist indexes with different specificities depending on the efficiency measure being analysed. The reasons for using a base period approach relate to the use of monthly data, and to the fact that technological change implies, in principle, longer periods than a month to take place. This means that we will not refer to technological change from month to month but to technological change happening in relation to a fixed period. Note that the use of monthly data may imply that frontier changes, measured by the technological change component of the Malmquist index, are due not only to technological reasons but also to seasonality. We will briefly discuss the issue of seasonality for each of the efficiency assessments undertaken, but note that this is not a major aim of our analysis over time.

The software used to produce all efficiency results shown in this Chapter was GAMS. As most of our models are not standard we needed to use a more flexible software that could handle every specificity of our models. GAMS proved very efficient and effective in this task.

10.2 Transactional Efficiency Assessment

The transactional efficiency assessment considers data from January 2002 to September 2002. The input-output variables used in this assessment are shown in Table 10.1.

Table 10.1: Inputs and Outputs used to assess transactional efficiency in month t

Inputs	Outputs
1. Number ETMs (ATMs + CATs) (t)	1. N. New registrations for multi-channel use (t)
2. Rent (t)	2. N. Transactions in CATs (t)
3. N. Clients not registered (t-1)	3.N. Deposits in ETMs (t)

The chosen inputs are intended to account for the resources that allow a bank branch to foster the use of alternative distribution channels, and the outputs are intended to capture the degree of usage of these channels. For details on the reasons behind our input-output choice, and also on the limitations of this choice please see Chapter 5.

Some statistics on the chosen inputs and outputs from January 2002 to September 2002 can be seen in Table D.1 in Appendix D. The values on this table show a stable behaviour of most variables over the period of analysis. It also shows that some outputs are zero for some branches. Since we use output VRS models this fact does not constitute a problem.

Transactional efficiency was assessed through the well known BCC model (Banker et al. (1984), defined in a VRS technology. The orientation we followed was towards output augmentation, since the objective is to increase transactions in alternative distribution channels given resources available at the branch. The BCC efficiency score is radial meaning that it does not account for all sources of inefficiency (see Chapter 2 for details). In order to consider all the sources of inefficiency we used an output oriented GDF measure. Recall that the GDF measure (introduced in Chapter 8) is defined as in (10.1), where $\theta_i = x_i^*/x_i$ and $\beta_r = y_r^*/y_r$.

$$\text{Geometric Distance Function (GDF)} = \frac{(\Pi_i \theta_i)^{1/m}}{(\Pi_r \beta_r)^{1/s}} \tag{10.1}$$

In our transactional efficiency assessment inputs are considered exogenously fixed and therefore each θ_i can be set equal to 1. As a result the output oriented GDF measure reduces to $1/(\Pi_r \beta_r)^{1/s}$, where output targets (y_r^*) are obtained directly from the BCC output oriented model.

In the transactional efficiency assessment we explored the existence of any influential or super-efficient observations in each month. When these observations were detected they were eliminated from the analysis so that final results were not distorted by the presence of influential bank branches. Existing procedures that can be used to detect influential or super-efficient observations have been detailed in Chapter 2. In the transactional efficiency assessment we detected influential observations by analysing super-efficiency scores and the number of times a unit appeared in the peer set of other units. Only one branch was considered to be super-efficient (B36) in January and February. This branch was eliminated from the technological set in these two months, since its presence was influencing the efficiency scores of the great majority of bank branches under analysis.

The presentation of results from the transactional efficiency assessment is divided into two parts. In the first part we present some detailed results for a given month, and in the second part we present some general results for all the months, and analyse the evolution of efficiency over time through Malmquist Indexes.

10.2.1 Detailed Results

The detailed results that are of interest for bank and branch managers involve the knowledge of bank branches' transactional efficiency scores, the knowledge of targets to be attained by bank branches, and the knowledge of peer or benchmark branches. Each of these results will be analysed next. Detailed results were produced for each bank branch in each month. Here, however, we discuss the results for only a few branches to illustrate the type of outcomes one can have for all branches. We have chosen results from the month of January 2002 for this illustrative purpose.

Efficiency Scores

The efficiency scores of all branches assessed in January 2002 are shown in Table D.2 in Appendix D. The VRS efficiency scores obtained are interpreted as usual radial efficiency scores. For example branch B15 has an efficiency score of 58.46%. Thus with its current levels of inputs, branch B15 is producing only 58.46% of the maximum possible outputs. In this sense the outputs of branch B15 should increase by about 171.07% (1/0.5846) so that this branch may reach the efficient frontier. The above efficiency score is radial in the sense that it is based on all outputs increasing by the same proportion towards the efficient frontier. If slacks exist, these are not accounted for in the 58.46% efficiency score for branch B15. In this case there are some slacks associated with the outputs 'registration' and 'transactions in CATs'. We can account for these slacks through the GDF output efficiency measure detailed earlier. The GDF output efficiency scores are also shown in Table D.2 in Appendix D. For branch B15 this score is 49%, which indicates that when slacks are accounted for the efficiency of branch B15 reduces by about 10 percentage points. This indicates that slacks represent a considerable source of inefficiency for this bank branch.

Consider another branch, B19, whose radial efficiency score is 42.97% and the GDF output efficiency score is 42.31%. In this case the percentage of inefficiency in the form of slacks is in fact negligible. On average, in the month of January output radial efficiency is 68.59% and the GDF output efficiency score is 64.16%. This means that there is a considerable degree of transactional inefficiency in the month of January, and most of this inefficiency is explained by radial inefficiency rather than by slacks. These account on average for about 4 percentage points inefficiency only.

Target Setting

Apart from efficiency scores, targets to be attained by inefficient bank branches are also of interest for bank and bank branch managers. Consider the two cases mentioned above of branches B15 and B19. The observed and target levels of these branches are shown in Table 10.2. We distinguish between radial targets (Rtargets) and non-radial targets (NRtargets). The differences between radial and non-radial targets reflect the value of slacks identified by the model. Targets in Table 10.2 contain information on how much each observed output should improve so that branches B15 and B19 can reach the efficient frontier. For the case of branch B15, the targets identified show that, given the number of clients, the number of ETMs, and the environmental conditions of this branch (reflected through the surrogate rent), it should in fact have a higher number of registrations on the multi-channel code, a higher number of deposits in ETMs, and a higher number of transactions in CATs. Branch B15 is not therefore exploiting at the maximum its inputs to generate the desired outputs. Non-radial targets for branch B15 show the existence of slacks on the outputs registrations and transactions in CATs. For branch B19 the interpretation of targets is similar, but in this case there are also input slacks. These input slacks should be seen as informational only, i.e. even if rent and clients had been smaller, B19 should still have been able to achieve the target outputs.

Table 10.2: Observed and Target Levels for Branches B15 and B19

		B15			B19		
		Observed	RTargets	NRTargets	Observed	RTargets	NRTargets
Inputs	Rent	4.591	4.591	4.591	4.779	4.779	4.539
	N. ETMs	3	3	3	2	2	2
	N. Clients	3810	3810	3810	4038	4038	3410.65
Outputs	Registr	10	17.107	22.377	13	30.2562	30.2562
	Deposits	1223	2092.186	2092.186	430	1000.782	1000.782
	TraCats	2905	4969.584	6449.044	1052	2448.425	2564.205

Peers or Benchmarks

The targets shown above were based on efficient branches to which each inefficient branch was compared. These efficient branches are the *peers* or *benchmarks* of the inefficient branch. Consider, for example, branch B19 whose peers are shown in Table 10.3. The peers of branch B19 are branches with, in general, lower inputs but higher outputs than B19. For example branch B42 has less of most inputs than branch B19 and produces much more

Table 10.3: Peers of Branch B19

			0.53	0.47
		B19	B32	B42
Inputs	Rent	4.779	5.574	3.353
	N. ETMs	2	2	2
	N. Clients	4038	3957	2785
Outputs	Registr	13	20	42
	Deposits	430	1590	326
	TraCats	1052	2594	2530

registrations in code multi-channel and more transactions in CATs. Branch B32 produces more of all outputs than branch B19 but one of its inputs (rent) is higher. Each of B32 and B42 can be seen by branch B19 as a role model branch whose transactional practices could be emulated.

10.2.2 Transactional Efficiency Results Over Time

More important than analysing efficiency in each month is to analyse its evolution over time. Indeed, for such a short period as a month we cannot state that a branch is efficient just because it was so in one month. Efficiency implies a consistent behaviour over a given period of analysis.

GDF output efficiency results computed independently for each month are shown in Table D.3 in Appendix D (branches that were deemed super-efficient are underlined). These results show that some branches were considered transactional efficient over the entire period of analysis. Such is for example the case of branches B4, B7, B36, B46, B48, and B54. Other bank branches presented a good overall behaviour in the same period, being efficient in most of the months under analysis but inefficient in some other months.

The analysis of efficiency scores observed in each period reveals little about the technological evolution of the variables used in the transactional efficiency assessment. In order to analyse this evolution we calculated Total Factor Productivity (TFP) change through the GDF approach as detailed in Chapter 9 and decomposed this TFP change into efficiency change (EFCH), technological change (THCH), and a residual component (RES) that is scale related. For reasons outlined earlier we used a base period approach to do that, where the month taken as a basis was January 2002.

To calculate the various components of TFP one needs to assess bank branches not only in relation to the frontier of the period to which data corresponds, but also in relation to

Table 10.4: Total Factor Productivity Change and its Components

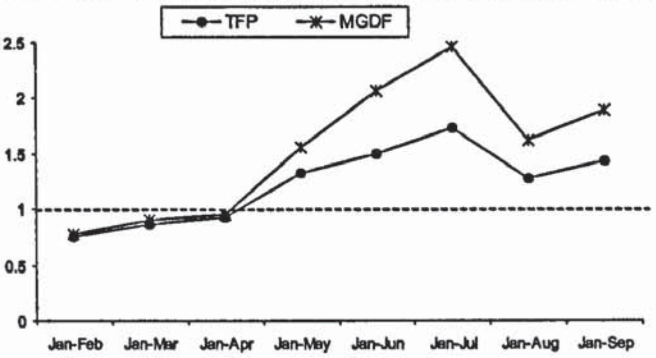
Period	GDF_b	GDF_t	$MGDF$	$EFCH$	$THCH$	RES	TFP
Jan-Feb	0.6272	0.7135	0.7764	1.1473	0.6565	1.0242	0.7510
Jan-Mar	0.6272	0.6124	0.900	1.0091	0.9159	1.0178	0.8699
Jan-Apr	0.6272	0.6590	0.9547	1.0849	0.9066	1.0223	0.9271
Jan-May	0.6272	0.6463	1.5636	1.0715	1.4572	0.9259	1.3182
Jan-Jun	0.6272	0.5985	2.0731	0.9894	2.1084	0.8017	1.4984
Jan-Jul	0.6272	0.7059	2.4573	1.2201	2.1458	0.7961	1.7331
Jan-Aug	0.6272	0.6357	1.6254	1.0813	1.5818	0.8738	1.2793
Jan-Sep	0.6272	0.6962	1.8964	1.1938	1.6296	0.8290	1.4408

the base period frontier. As we used BCC models in our assessments the models assessing units in relation to a different time period frontier may not have a solution. The number of branches for which a Malmquist index could not be calculated was not, however, very significant in each month (the highest number of branches in this situation was two per month), and therefore the non-inclusion of these branches in our computations did not seriously distort the average results that we present in Table 10.4. In this Table we show TFP average values and its components, and also the average efficiency scores obtained in each month (GDF_b, corresponding to the average efficiency score of the base period, and GDF_t corresponding to the average GDF efficiency score of month t).

These values show that July was the best month in terms of productivity growth and February was the worst month when compared to January 2002. The two main components of TFP are the Malmquist GDF index ($MGDF$) and the Residual (RES) component (see Chapter 9, expression (9.10)). The evolution of the former component is very much alike the evolution of TFP as can be seen in Figure 10.1. The difference between TFP and MGDF is the residual effect, and therefore in Figure 10.1 we can see that this effect is almost negligible in the first three months but then it begins to assume some importance in the remaining months. Note that TFP is increasing in relation to January for almost all the months under analysis except August (in this month TFP change decreases in relation to the July TFP change, but it is still above one indicating a TFP increase from January to August 2002).

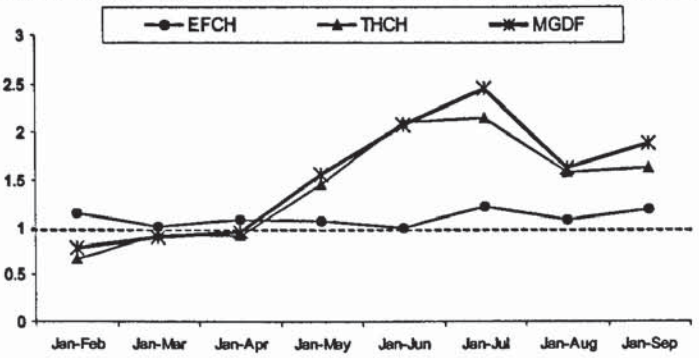
The two components of the $MGDF$ are technological change (THCH) and efficiency change (EFCH). In February, March and April there was technological regress in relation to January, but afterwards there is a remarkable technological progress. This means that output levels on the frontier increased considerably after April when compared with output levels in January. Note that the component that most explains the MGDF component of

Figure 10.1: Total Factor Productivity and Malmquist GDF Index



TFP is technological change rather than efficiency change. This is clear in Figure 10.2, where we show each component of the MGDF.

Figure 10.2: Efficiency Change and Technological Change



Efficiency change values are very close to 1 and therefore they do not seem to affect particularly the MGDF component and therefore the TFP change. The productivity improvement of 1.4408 from January to September 2002 in the transactional efficiency assessment is therefore mostly dictated by changes in the frontier rather than by changes in efficiency. This means that bank branches delineating the efficient frontier of September showed on average higher outputs (registrations, transactions in CATs and deposits in ETMs) and/or lower inputs (rent, N. ETMs, and N. clients) than those delineating the January frontier.

Seasonal effects are not evident in the decomposition of TFP change for the transactional efficiency assessment. Note however, that the peak of technological change around Summer months may be a result of seasonal effects that moved the frontier upwards. As we do not have data for more than one set of Summer months we cannot ascertain anything regarding possible seasonal effects on the transactional efficiency assessment.

10.3 Operational Efficiency Assessment

The operational efficiency assessment considers data from March 2001 to September 2002. The input-output set used in this assessment is shown in Table 10.5, where (t) denotes time period and the Greek Δ denotes change in values between the start and the end of period t. The inputs we have used reflect the main operational resource of bank

Table 10.5: Inputs and Outputs used to assess operational efficiency in month t

Inputs	Outputs
1. Number Staff (t)	1. Δ Number of Clients (t)
2. Rent (t)	2. Δ Value Current Accounts (t)
	3. Δ Value Other Resources (t)
	4. Δ Value Titles Deposited (t)
	5. Δ Value Credit Over Bank (t)
	6. Δ Value Credit Associates (t)
	7. Number Transactions (t)

branches (staff) and its environmental conditions (rent). The outputs we have chosen are intended to reflect the main operational objectives of bank branches (to increase the customer base, to increase sales of the various products the branch has to offer, and to serve clients). For further details on the reasons behind our input-output choice and limitations concerning this choice the reader is referred to Chapter 5.

As specified in Table 10.5 some outputs may be negative and therefore operational efficiency is measured through the RDM models described in Chapter 6, which can handle negative data. The RDM models are used here with orientation towards output enhancements, though one of the outputs considered is non-discretionary. The non-discretionary of transactions is treated according to the procedure of Banker and Morey (1986a) as detailed in Chapter 6.

Before applying the RDM models we checked our data for 'super-efficiency' or influential units. Although there are no super-efficiency models (Andersen and Petersen, 1993) that can be applied directly to the case of negative data one can use the additive model to perform a diagnosis of super-efficiency. This can be done by analysing units that eventually dominate the reference set and re-running the models without such units. Common sense plays, necessarily, an important role in determining which units are influential, while at the same time complementary information from the bank should also be sought in this process. In our case, for example, in some months some bank branches closed and their business was transferred to other branches in the sample. This means that the latter branches suffered an abnormal increase in their business in a particular month, which was not a result of an improvement in their efficiency but a result of uncontrollable factors. Branches in this situation were eliminated from the reference set in the month where this situation happened.

The results on the operational efficiency assessment will be presented first in terms of general outputs from the RDM efficiency assessments and then we will analyse the evolution of efficiency over time.

10.3.1 Detailed Results on Operational Efficiency

Detailed results were produced for each month under analysis. Here, however, we discuss the results for only a few branches to illustrate the type of results one can have for all branches. We have chosen results from April 2001 for our illustrative purpose. Descriptive statistics on the operational efficiency inputs and outputs are shown in Table E.1 in Appendix E from March 2001 to September 2002.

The detailed operational efficiency results that are of interest to managers involve the knowledge of efficiency scores through which bank branches can be ranked and classified into efficient or inefficient, the identification of targets to be attained by inefficient bank branches, and also the identification of peer or benchmark branches whose operational practices should be emulated by inefficient branches.

Efficiency Scores

The application of the RDM model resulted in a set of efficiency measures that allowed us to assess how far each bank branch is from its Pareto-efficient targets, and how branches compare between each other in terms of this distance. The detailed RDM efficiency scores for each branch in the month of April 2001 are shown in Table E.2 in Appendix E. We produced two different kinds of efficiency scores based on the RDM model. One is based directly on the results from the RDM model and consists on the values of $1-\beta$. Recall that β is an inefficiency measure and therefore $1-\beta$ is an efficiency measure. In the RDM model the $1-\beta$ values are interpreted as the radial distance from each bank branch to its targets. For example, branch B11 has a $1-\beta$ value equal to 72.74% and branch B12 has $1-\beta$ equal to 90.61%. This means that targets of branch B12 are radially closer from its observed levels than the targets of branch B11. This radial distance does not reflect, however, the potential for non-radial improvement in some outputs (note that improvement in inputs is not sought in our case since we use output oriented measures). For this reason, we use a ratio of norms ($||I - T||/||I - O||$, where I stands for ideal point, T for target levels and O for observed levels) as detailed in Chapter 6 to account for all the sources of inefficiency associated to outputs (the ratio of norms is computed based on normalised data and on discretionary outputs only). According to the ratio of norms unit B11 has an efficiency score of 65.39 % and unit B12 has an efficiency score of 88.15%. When all the sources of inefficiency are accounted for, branch B12 still ranks better than unit B11, but in both cases the measure of efficiency decreased meaning that there were some slacks not accounted for by the radial efficiency measures.

For the month of April 2001 the RDM average $1 - \beta$ is equal to 89.1%, and the average ratio of norms is 85.4%. The closeness between these values means that slacks were not on average an important source of inefficiency in this month.

In terms of rankings the correlation between the rank based on the RDM efficiency score and on the ratio of norms is very high (0.9393) meaning that these efficiency scores result in very similar ranks. Nevertheless there is a large difference in rank for some branches. For example, the highest difference happens for branch B43, which is ranked 19 under the RDM efficiency score and 39 when slacks are taken into account in the ratio of norms.

Target Setting

For the bank it is not only important to know how each unit compares in terms of their distance to targets, but also to know what those targets are. A bank branch can seek targets that stress attaining more on those factors where performance is poorest at present, or on those where performance is best at observed levels. As we saw in Chapter 6 these two aims are served respectively by the RDM and the IRDM models (recall that the IRDM model is similar to the RDM model except on the directional vector, which is the inverse of the range of possible improvement rather than the range itself). We produce here both sets of targets to illustrate the difference. Table 10.6 shows the observed and target levels of units B8, B15 and B19 in April 2001.

The advantage of using both the RDM and IRDM models is that we can provide alternative targets that represent different routes that the bank branch can choose in order to become efficient. The RDM and IRDM procedures clearly give different priorities on improving different variables, with RDM targets being more demanding in certain variables, while the IRDM targets being more demanding in others. Take for example branch B19,

Table 10.6: Target Levels for Some Units in April

		B8			B15			B19		
		Obs	RDM	IRDM	Obs	RDM	IRDM	Obs	RDM	IRDM
Inputs	Rent	2.36	2.15	2.17	4.4	3.34	3.51	4.58	2.13	2.28
	Staff	4	4	4	7	5.52	5.73	5	5	5
Outputs	ΔCli	-4	34.71	37.91	-32	38.46	24.12	-32	43.68	36.87
	ΔCurac	-102.428	-25.53	-40.76	-90.282	36.89	0.59	-44.029	83.18	75.73
	ΔOthre	127.443	268.29	281.35	259.335	550.13	623.24	174.594	433.03	457.17
	ΔTdep	-49.506	-30.71	-39.17	-102.917	-61.19	-91.57	-71.707	-33.22	-55.49
	Δ Credb	128.416	298.44	311.55	188.037	403.85	466.97	-18.071	257.98	271.95
	ΔCredas	5.183	52.90	50.84	66.634	101.43	128.12	10.679	62.40	73.29
	Trans	2983	2983	2983	3441	4958.28	4983.33	3834	3834	3834

Table 10.7: Output Improvements for Branch B19

	ΔCli	ΔCurac	ΔOthre	ΔTdep	ΔCredb	ΔCredas
Range	1.09	1.076	0.837	1.608	1.014	0.96
RDM	75.675	127.211	258.439	38.49	276.048	51.723
IRDM	68.873	119.755	282.577	16.218	290.017	62.606

whose output range of improvement (normalised by the maximum output so that different units of measurement do not distort our interpretations) is (1.09, 1.076, 0.837, 1.608, 1.014, 0.96) (see Table 10.7). The highest value of this range occurs for the fourth output (ΔTdep), which means that the RDM model will give priority to improving this output, while the IRDM will give priority to improving the third output (ΔOthre) and eventually the last output (ΔCredas). Note that the output improvements (difference between target and observed values) corresponding to the RDM and IRDM targets of unit B19, shown in Table 10.7, clearly confirm that factors with a higher range improve more under the RDM, and factors with a lower range improve more under the IRDM model.

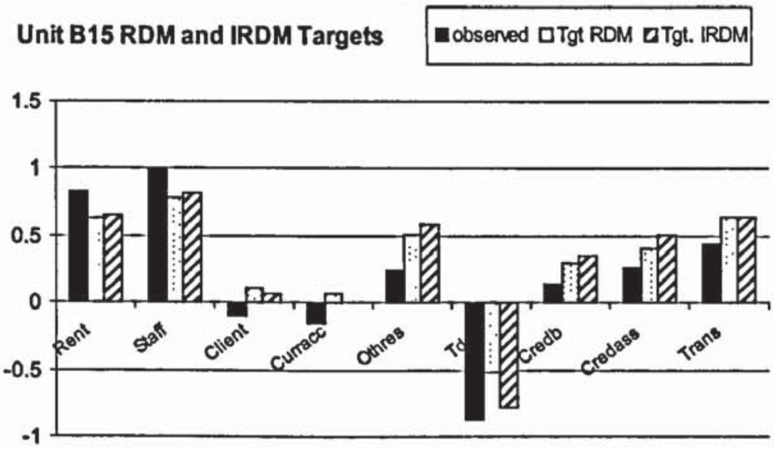
The improvements in Table 10.7 are not units invariant and therefore it is not easy, by simply inspecting this table, to ascertain whereas the RDM or the IRDM targets are closer to observed values. The efficiency scores as measured by the ratio of norms, however, show that IRDM targets are closer than RDM targets for branch B19 (IRDM distance is 82.63% and RDM distance is 78.78%), meaning that IRDM targets require a smaller effort from branch B19 to reach 100% efficiency. Obviously if unit B19 has particular difficulty in selling 'other resources' items and foresees that it cannot improve it as much as given by the IRDM targets it has the alternative of using targets given by the RDM model.

The ratio of norms efficiency score was calculated for all branches considering both their RDM and IRDM targets. On average the ratio of norms for the RDM model is 85.44% while the average ratio of norms for the IRDM model is 88.67%. This means that on average IRDM targets are in fact closer from observed levels than RDM targets.

The targets in Table 10.6 may reveal additional inefficiencies apart from those associated with discretionary outputs. These should be used as informational only. For example we can say that branch B15 could have achieved the output targets even if it had less staff, which suggests that this bank branch is over-staffed. At the same time branch B15 could have achieved the same targets even if it had performed more transactions which seems to indicate that staff have some free time that could be used on other activities.

An alternative representation of the targets in Table 10.6 is through a graph as that shown in Figure 10.3. In this Figure targets and observed values are all normalised by the maximum value, and the objective is simply to provide an easy tool from which managers can immediately pick up what is going on.

Figure 10.3: Graphical Representation of Targets



Peers or Benchmarks

As usual in DEA analyses apart from targets, we can also provide inefficient units with peer or reference units to which they can compare themselves. For the case of branch B15, Table 10.8 presents the reference units as given by the RDM and IRDM models. Although here the RDM and IRDM models yielded the same peers, this is not the case in general.

The present peers are benchmarks that can be used by branch B15 as references of performance in the month of April. These benchmarks explain why branch B15 is inefficient and how targets for this branch have been constructed. All of the benchmarks perform better than B15 in most respects, while having less favourable conditions in terms of the

Table 10.8: Peers of Branch B15

		RDM	0.1642	0.0158	0.2049	0.0799	0.5353
		IRDM	0.1046	0.1764	0.0939	0.0853	0.5397
		B15	B58	B41	B38	B16	B9
Inputs	Rent	4.40	3.352	4.778	3.093	4.025	2.82
	Staff	7	5	7	5	7	7
Outputs	ΔCli	-32	-21	-8	-29	18	341
	ΔCurac	-90.282	-124.5	-48.503	582.99	149.43	-87.798
	ΔOthre	259.335	598.782	807.783	-108.349	732.984	1070.42
	ΔTdcp	-102.917	118.011	-494.942	-80.002	-372.048	-381.042
	ΔCredb	188.037	338.803	1330.548	211.575	626.585	383.77
	ΔCredas	66.634	130.86	-18.725	97.964	250.85	53.91
	Trans	3441	3846	5508	5322	6403	7724

inputs (rent and staff) and in terms of the non-discretionary output: transactions. Note that branch B9 dominates the peer set both for the RDM and IRDM models. B9 has the same staff as B15, it is located in a less favourable region (rent is lower), and it performs many more transactions than branch B15. Even so, branch B9 performs better than branch B15 in most outputs. Branch B9 can therefore be looked at by branch B15 as a benchmark unit whose operational practices could be emulated. Depending on the chosen targets (RDM or IRDM) benchmarks may be different. For example for IRDM targets the main benchmarks for branch B15 are B9, B41, and B58, while for RDM targets the main benchmarks are B9, B38, and B58. Clearly these benchmarks represent different output mixes from which B15 can choose as it prefers.

10.3.2 Operational Efficiency Results Over Time

In Table E.3 in Appendix E we show detailed results on operational efficiency for each month under analysis, where branches are sorted by their average over time efficiency. Branches that were excluded from the analysis in each month, because they were deemed super-efficient, are underlined in this table. In order to consider a bank branch efficient, it should show a consistent efficient behaviour over the period of analysis. Such are branches B4 and B9, and also branches B2, and B28. The two latter bank branches closed, however, in October 2001[1]. Other bank branches, showed 100% efficiency in most of the months under analysis and therefore can be said to have good operational efficiency (e.g. B33, B54,

[1]The reasons for the bank closing some bank branches was not directly related to the performance of these branches. In most cases the decision had to do with more than one bank branch operating very close to each other, and the bank decided to close on of these branches (for example the smaller one since the bank branch that remained open must absorb the customers of the closed branch).

B27, B46, etc.).

The values shown in Table E.3 relate to the assessment of efficiency in each month separately. We also used RDM based Malmquist indexes to analyse efficiency and technological change over time (see details in Chapter 6). For the reasons outlined in the introduction to this Chapter, we used a base period approach where March 2001 was used as the base technology. For this purpose it was necessary to define an ideal point to be used in the assessment of units against two technologies (the base period technology and the current period technology). Such an ideal point was renewed every month since we considered it to be the maximum output levels in each of the two periods under analysis (base period and current period). Note that our application of the RDM model is oriented towards output enhancements and therefore the ideal point to be considered only needs to account for outputs. For details please refer to section 6.6 of Chapter 6.

A summary of the results obtained from each month is presented in Table 10.9. The

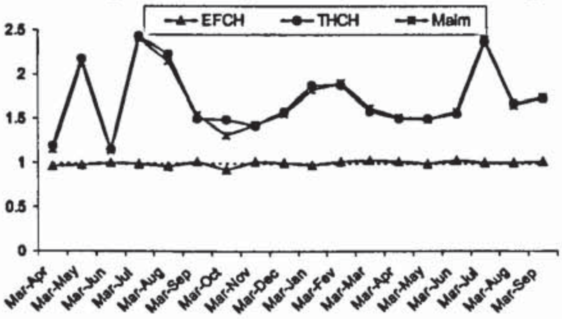
Table 10.9: Base Period Results

Period	$\gamma^b(y_b, x_b)$	$\gamma^t(y_t, x_t)$	N. Inf.	EFCH	THCH	Malm.
Mar01-Apr01	0.9803	0.943	5	0.9624	1.1894	1.1478
Mar01-May01	0.983	0.9505	11	0.9673	2.1727	2.1336
Mar01-Jun01	0.9819	0.9757	7	0.9942	1.1425	1.1343
Mar01-Jul01	0.983	0.9712	15	0.989	2.434	2.4036
Mar01-Aug01	0.9781	0.9386	9	0.9608	2.2348	2.1474
Mar01-Sep01	0.9598	0.9727	6	1.0193	1.4987	1.545
Mar01-Oct01	0.9759	0.8945	14	0.9177	1.4894	1.3056
Mar01-Nov01	0.9595	0.9727	10	1.0185	1.4159	1.4349
Mar01-Dec01	0.9783	0.9723	26	0.9953	1.5787	1.5592
Mar01-Jan02	0.9714	0.9387	4	0.9689	1.8798	1.8182
Mar01-Feb02	0.9585	0.9664	8	1.0136	1.8769	1.9083
Mar01-Mar02	0.958	0.9843	5	1.033	1.5769	1.6172
Mar01-Apr02	0.9638	0.9761	12	1.0165	1.5023	1.5151
Mar01-May02	0.9893	0.9782	26	0.9891	1.4936	1.4877
Mar01-Jun02	0.956	0.9729	20	1.0241	1.5475	1.5853
Mar01-Jul02	0.9612	0.9521	33	0.9951	2.3694	2.3873
Mar01-Aug02	0.973	0.9666	22	0.9957	1.6722	1.6549
Mar01-Sep02	0.955	0.9609	32	1.0113	1.7356	1.7371

column named 'N. Inf.' in this Table indicate the number of infeasible models that were found in assessing data of branches in t in relation to the frontier of the base period (for details see section 6.6 of Chapter 6). Bank branches with infeasible RDM models are located above the base period frontier, indicating that the technological change component (THCH) in Table 10.9 in under-estimated. The RDM $1-\beta$ efficiency of each period is given by $\gamma^t(y_t, x_t)$ for the period t assessment in relation to the t frontier, and $\gamma^b(y_b, x_b)$ for the base period assessment in relation to the base period frontier. Note that $\gamma^b(y_b, x_b)$ is different for each line of Table 10.9 because the ideal point against which efficiency is measured varies from month to month.

The values in Table 10.9 show that there was always technological progress in relation to March 2001, though in some months this progress was more intense than in others. The months with higher progress in relation to March 2001 are May, July, and August 2001 and July 2002. In Figure 10.4 we plot the Mamquist index and its technological change and efficiency change components. Technological change seems to be the main factor explaining changes in the Malmquist index, since efficiency change is mostly around one. Note, however, that technological change seems to be confounded in this case with seasonal effects that push the frontier upwards in Summer months (July 2001 and July 2002 are the months that exhibit highest progress). Therefore, what we are calling technological change in Table 10.9 means in fact a frontier movement that may be attributable both to changes in the technology and to seasonality.

Figure 10.4: Technological Change and Efficiency Change Components



In order to analyse the extent to which there are seasonal effects one can compute technological change values for data one year apart. For that purpose we can use simple algebra and calculate from Table 10.9 the technological change values shown in Table 10.10, where, for example, the value from July 2001 to July 2002 was obtained from the ratio between THCH(Mar01, Jul02) and THCH(Mar01, Jul01). Values in Table 10.10 show that technological change in July 2002 when measured in relation to July 2001 is very close to one, denoting no technological progress between these two months. Note that technological change values in Table 10.10 were calculated algebraically (through the ratios of technological change in relation to the base month of March 2001) and not by direct application of the RDM to data one year apart. This latter way of calculating technological

Table 10.10: Technological Change without Seasonality

Period	THCH
Mar01-Mar02	1.5769
Apr01-Apr02	1.2631
May01-May02	0.6874
Jun01-Jun02	1.3545
Jul01-Jul02	0.9735
Aug01-Aug02	0.7483
Sep01-Sep02	1.1581

change does not necessarily lead to the same results since the Malmquist index based on the RDM does not satisfy the circularity property. Our purpose here is not, however, to detail too much on seasonal effects but only to acknowledge their possible existence and influence on frontier movements.

Ignoring monthly fluctuations of efficiency change and technological change we can say that in this period of about one year and a half, there was an important amount of technological progress measured by an index of about 1.7 (see Table 10.9). This means that the frontier moved upwards, but on average bank branches kept the pace with these frontier movements, since the average efficiency of bank branches remained more or less unchangeable during the whole period of analysis.

10.4 Profit Efficiency Assessment[2]

The profit efficiency assessment also considers data from March 2001 to September 2002. The input-output variables used in this assessment are shown in Table 10.11 (see Table F.1 in Appendix F for descriptive statistics on the chosen inputs and outputs over the period of analysis). These variables are consistent with the intermediation approach of

Table 10.11: Inputs and Outputs used to assess profit efficiency

Inputs	Outputs
Number of Staff [Staff]	Value Current Accounts [Curracc]
Supply costs [Supplycost]	Value Other Resources [Othress]
	Value Credit by Bank [Credb]
	Value Credit by Associates [Credass]

bank branches' activities as discussed in Chapter 3. For a more detailed discussion on the

[2]**Part of the results presented in this section will appear in Portela and Thanassoulis, *Profitability of a sample of Portuguese bank branches and its decomposition into technical and allocative components*, European Journal of Operational Research.**

reasons and limitations behind our choice of variables please see Chapter 5.

Apart from the 'quantity' variables specified in Table 10.11, price data were also available for staff (average salaries) and for all outputs. We used the variables in Table 10.11 to compute two types of efficiency: technical and overall profit efficiency. For the technical profit efficiency measurement we used the variables as specified in Table 10.11. For calculating overall profit efficiency we used the 'quantity' data specified in Table 10.11 plus price information for all the variables in this table, except for supply costs (see section 8.5 of Chapter 8). Recall that output prices are average net interest rates and not individual interest rates per bank branch (see Chapter 5).

Overall profit efficiency was calculated and decomposed into its technical and allocative components using the Geometric Distance Function (GDF) procedure detailed in Chapter 8. These computations were performed assuming a long run and a short run settings. The detailed results, presented in the next sections, consider both these situations.

10.4.1 Long Run Detailed Results

Detailed results were produced for every month under analysis, but for illustrative purposes we discuss here only results obtained for April 2001.

Recall that the GDF measure requires the computation of technical efficient and maximum profit targets. These targets reflect alternative ways a bank branch can follow to increase its profits. Technical efficient targets, for example, always imply increasing outputs and reducing inputs in movements of inefficient units towards the frontier, which obviously increase their profit. Profit targets on the other hand focus on further enhancements that could be achieved through an adequate management of the bank branch's product mix given the factor prices they face. Technical and profit targets are calculated here first without any constraints (long run approach), and then with certain constraints to targets such that these are in fact possible to achieve in the short run. As the identification of targets is done prior to the measurement of efficiency we illustrate first the computation and interpretation of targets, and then illustrate the calculation of various efficiency measures.

Long Run Targets

Long run technical efficient targets are found through the closest target (CT) procedure detailed in Chapter 7. Maximum profit targets are calculated based on the maximum profit model (8.2) shown in Chapter 8. The calculation of these targets allows the computation of overall profit efficiency and technical profit efficiency. The allocative profit efficiency is obtained by decomposition, and it can be further decomposed into scale and mix effects. Scale adjusted targets can also be computed by applying model (8.7) detailed in Chapter 8.

Take for example unit B8 whose various types of targets obtained in the long run analysis are shown in Table 10.12. We consider in this table two types of technical efficient targets

Table 10.12: Long Run Targets for Unit B8

	Inputs		Outputs				
	Staff	Supplycost	Curracc.	Othrcss.	Credb	Credass.	GDF
B8 - Obs.	4	13.2181	2009.8513	4751.0899	4986.2581	248.7056	
Additive Tgt	3.2603	13.2181	2463.4065	7856.6609	7870.4677	2100.876	39.60%
Close Tgt	3.3481	10.7413	2009.8513	4902.2491	4986.2581	248.7056	81.83%
Max Profit Tgt	7	26.8104	7361.818	20266.4129	21512.1108	2306.9752	37.68%
Scale adj.	8.357	26.8104	7361.818	17956.2865	18264	910.976	

to show that the CT procedure is one out of a number of ways through which technical efficient targets can be computed. The alternative we used here is the additive model. Note that the main difference between CT targets and targets from the additive model is that the former are easier to achieve and imply smaller adjustments on the current level of inputs and outputs of branch B8. Smaller adjustments are reflected in a higher GDF efficiency score for the CT target (81.83%) than for the target resulting from the additive model (39.6%)[3].

The maximum profit target of unit B8 corresponds to the observed input/output levels of unit B16. In fact, the long run profit maximising model rendered unit B16 as the single maximum profit target for all bank branches under analysis. This is in principle a result of very similar prices across all bank branches, which does not allow for much discrimination in terms of profit maximising units. In order to check the sensitivity of the maximum profit unit to changes in factor prices, we simulated price changes by adding to the original output prices a random number varying between 0 and 0.02. Note that our prices are interest rates and vary between 0.0059 for credit by the bank and 0.0419 for current accounts[4]. The results from our simulated price data are very similar to those

[3]For the overall set of 57 bank branches the average GDF found for the unit's invariant additive model is 67.98% (with a minimum of 10.25%), while this average is 82.99% (with a minimum of 37.91%) for the CT procedure. This indicates that the CT in fact provides closer targets which are reflected in higher efficiency scores.

[4]The price of credit is lower than the price of current accounts, because the interest rate of credit includes a risk factor that does not exist for current accounts.

with constant prices. With simulated prices unit B17 and B16 appear most of the times as the profit maximising branches. In some attempts only branch B17 appeared and in other attempts only branch B16 appeared as profit maximising. This fact seems to suggest that our results are not very sensitive to the fact that equal prices across branches were used for outputs.

The scale adjusted targets in Table 10.12 reflect the change in scale that is needed to move from the technical efficient target to the maximum profit target. In the case of branch B8 scale changes in inputs are equal to 2.496 and scale changes in outputs are equal to 3.663. This means that the adjustments required in branch B8 to go from the technical efficient target to the maximum profit target imply changing outputs more than proportionally compared to changing inputs. This suggests increasing returns to scale movements towards maximum profit. The distance between maximum profit targets and scale adjusted targets reflects mix changes that are a component of allocative efficiency.

Long Run Efficiency Scores

The GDF is applied *a posteriori* to express the distance between an observation and its target. In this sense applying the GDF between the observed input/output levels of branch B8 and its CT technical efficient targets renders a GDF score of 81.83%. Applying the GDF to measure the distance between the observed point and the maximum profit point (unit B16) renders a GDF overall profit efficiency score of 37.68%. The allocative GDF efficiency measure can be computed by applying the GDF to measure the distance between CT technical efficient targets and maximum profit targets. Alternatively the value of the GDF allocative efficiency can be found by decomposition since overall profit efficiency equals the product of technical efficiency and allocative efficiency. Applying such a decomposition to branch B8, renders an allocative efficiency score of $\frac{37.68\%}{81.83\%} = 46.042\%$. This allocative measure of efficiency can be further decomposed into a mix and scale effect. The scale effect equals the ratio of the proportional adjustments in inputs and the proportional adjustments in outputs, i.e. $\frac{2.496}{3.663} = 0.6814$. Applying the GDF to measure the distance between the scale adjusted target and the maximum profit target results in a mix effect of 0.6757. Note that a scale effect lower than 1 indicates increasing returns to scale movements, while a scale effect larger than 1 indicates decreasing returns to scale movements (see Chapter 8 for details). The product of all efficiency scores identified before is the overall profit efficiency measure of unit B8, that is, $0.8183 \times 0.6814 \times 0.6757 = 0.3768 \Leftrightarrow$ technical efficiency $\times$ scale effect × mix effect = technical efficiency × allocative efficiency = overall profit efficiency.

As shown in Chapter 8 allocative and overall profit efficiency can be higher than 1. Values of profit or allocative efficiency greater than 1 indicate movements that are not advisable from a purely technological perspective but are so from a profit perspective. As can be seen in Table F.2 in Appendix F, which summarises results for all the branches under analysis for April 2001, profit and allocative efficiencies are higher than 1 for unit B28. If we calculate for branch B28 an aggregate measure of productivity through the GDF $((\Pi_r \frac{y_r}{staff} \times \Pi_r \frac{y_r}{supplycost})^{\frac{1}{s \times m}})$, this is 814.076 at the observed point (which coincides with the technical efficient point), and 677.1288 at the maximum profit point. This means that a movement from the technical efficient point to the maximum profit point implies a decrease in aggregate productivity, and as such this movement is not advisable under a technical perspective. In fact the aggregate productivity at the observed point is $\frac{814.076}{677.1288} = 1.202$ times higher than the productivity at the maximum profit point. Values greater than 1 should therefore be interpreted carefully because the trade-offs implicit in these values might be of such an order that are not in fact advisable or possible.

Results in Table F.2 in Appendix F show that average profit inefficiency of April 2001 is 39.43% (measured as the absolute deviation of profit GDF from 1). This value is mostly explained by allocative inefficiencies (average allocative inefficiency is 27.39%) and less by technical inefficiencies (average technical inefficiency is 17.01%). This means that though branches are relatively close to the technical efficient frontier, this frontier is far away from the maximum profit plane (that passes through unit B16) resulting in high allocative inefficiencies. The allocative (in)efficiency is composed of mix and scale effects. The average scale effect is 1.0243 which indicates that on average branches increased more inputs than outputs in moving from its technical efficient projection to the maximum profit point. This fact indicates average decreasing returns to scale in such movements, but these are not very marked as the scale effect value is very close to 1. The mix effect on the other hand seems to be the main reason that the allocative efficiency averages 71.5%.

Long Run Efficient Peers

For each inefficient branch our analysis identifies efficient bank branches whose performance they could emulate to improve their long run profitability. The maximum profit model identified a single branch as the efficient peer of all branches. As far as technical profit efficient targets are concerned different peers were identified for different inefficient

Table 10.13: Peers of Branch B8

			0.4224	0.303	0.0225	0.2521
		B8	B2	B33	B37	B54
Inputs	staff	4	3	4	5	3
	supplycost	13.2181	9.2776	14.5749	14.8342	8.2202
Outputs	curracc	2009.8513	1161.7352	4127.1885	4012.7493	706.8814
	othress	4751.0899	2464.5953	9964.8148	7458.1858	2673.2225
	credb	4986.2581	5776.2692	5697.5938	14226.0901	1981.2552
	credass	248.7056	173.3622	422.5467	1414.9699	61.7412
Profit	151.634	97.3051	309.5077	340.8308	57.7969	

branches. We use branch B8 to illustrate the results here (see Table 10.13).

The efficient peers of branch B8 have in general lower inputs and higher outputs than B8. This fact explains the inefficiency of branch B8. Consider for example branch B33 that has the same number of staff of branch B8 and higher supply costs. B33 is able to achieve much more on all outputs than branch B8, achieving thus a higher profit. Branch B8 would, therefore, benefit by emulating branch B33.

Profit Gains

The above analysis considers only changes in inputs and outputs from one point to another measured through the GDF. Complementing the GDF results with profit ratios or differences provides interesting insights. Consider for example branch B17. Its actual profit (which is also technical efficient profit) is about 663.9 thousand Euros, and its maximum profit is very close to it: about 677 thousand Euros (see Table F.2 in Appendix F). This seems to suggest that the profit efficiency of this branch is very high as it is very close to maximum profit. However, analysing the GDF we see that its value is small (67.76%). This basically means that in order to move from its actual position to the maximum profit point branch B17 needs to undertake considerable changes in the mix of its inputs and/or outputs. The main trade-off asked here to branch B17 is a large reduction (in about 32%) on the value of other resources (on which this branch performs better than B16) and a high increase (in about 350%) on the value of credit over associates. Such a trade-off might be questioned by branches whose profit is already close to maximum profit.

The overall gains in profit obtained by moving from the observed point to the maximum profit point is a sum of the profit gained by moving from observed points to technical efficient points (technical profit gain) and the profit gained by moving from technically efficient points to maximum profit points (allocative profit gain). In our case branches gain on average 41.5 thousand Euros by eliminating technical inefficiency and gain on average 294.867 thousand Euros by eliminating allocative inefficiency. Overall profit gains (on average 336.34 thousand Euros) are therefore mainly attributed to allocative movements rather than technical movements. This means that while the increase in profits obtained from the elimination of technical inefficiency might be possible to achieve in the short run, those accruing from the elimination of allocative inefficiency might be unrealistic in the short run as they imply a large change in inputs and outputs of bank branches. Changing the staff from 3 to 7, for example, might be difficult given the dimension of the bank branch and the benefits accruing from this change might not correspond to the expected due to environmental or market constraints. This is the reason why we analyse here not only long run models but also short run models where more realistic targets are proposed to bank branches.

10.4.2 Short Run Detailed Results

As seen previously, long run targets may be unrealistic and not achievable in the short run. As the bank was interested in knowing targets to achieve in the short run we added some constraints to the profit model as explained in section 8.6 of Chapter 8. Specifically we restricted inputs not to change by more than 30% from their observed levels. Bounds on outputs were also imposed in order to prevent big changes in the output levels and their mix. We constrained outputs to change by no more than 30% from their observed level. The type of results that can be obtained from a constrained profit model are shown in the next sections, where the month of April 2001 is once again used for illustrative purposes.

Short Run Targets and Efficiency Scores

The main results obtained from the short run profit efficiency assessment are a set of efficiency scores and short run targets to be attained by inefficient branches. For illustrating these results consider branch B8, whose targets obtained from the constrained models are shown in Table 10.14. In this Table we also show the long run targets for comparative purposes.

Long run technical efficient targets were found following the closest target procedure and therefore we should not expect big differences between short run constrained technical efficient targets and long run technical efficient targets. The biggest differences between short and long run targets happen for maximum profit targets, and this is clear for branch

Table 10.14: Long Run and Short Run Targets for Branch B8 in April 2001

	Inputs		Outputs				
	Staff	Supplycost	Curracc.	Othress.	Credb	Credass.	GDF
Long Run Targets							
B8 - Obs.	4	13.2181	2009.8513	4751.0899	4986.2581	248.7056	
Close Tgt	3.3481	10.7413	2009.8513	4902.2491	4986.2581	248.7056	81.83%
Max Profit Tgt	7	26.8104	7361.818	20266.4129	21512.1108	2306.9752	37.68%
Scale adj.	8.357	26.8104	7361.818	17956.2865	18264	910.976	
Short Run Targets							
B8 - Obs.	4	13.2181	2009.8513	4751.0899	4986.2581	248.7056	
Technical Tgt	3.5989	12.1302	2612.8067	6176.4169	6482.1355	323.3173	69.897%
Max Profit Tgt	3.3663	12.2623	2612.8067	6176.4169	6482.1355	323.3173	67.97%
Scale adj.	3.6381	12.2623	2612.8067	6176.4169	6482.1355	323.3173	

B8 in Table 10.14. For example, after reaching the technological constrained frontier B8 only needs to change its inputs to become a profit maximising unit. Note how these changes contrast with the long run case where differences between technical efficient targets and maximum profit targets are much higher. As a result, short run allocative efficiency is higher than the long run allocative efficiency for B8. Allocative efficiency now equals 97.24% (=0.6797/0.69897), as opposed to a value of 46.04% for the unconstrained long run profit model. Note that trade-offs within outputs are much lower in the short run than in the long run because of the constraints imposed.

As far as the constrained maximum profit model is concerned all the branches have at least one of the ratios between maximum profit target outputs and observed outputs equal to the upper bound of 1.3. This means that if branches are to go beyond the constrained profit they need to change at least one output by more than 30%. The bounds on inputs, in contrast, do not have a big impact on both the maximum profit and the technical efficiency models, as most of these bounds are not binding.

The detailed short run results for the month of April 2001 are presented in Table F.3 in Appendix F. These results show an average profit inefficiency that is lower than before. It is now 20.35% while in the unconstrained long run model it was 39.43%. The main differences between this model and the long run model concern the factors that most explain the profit inefficiency. In the short run constrained model most of the profit inefficiency is explained by technical inefficiency, which is on average 15.48%. Allocative effects are very small as confirmed by an average allocative inefficiency of 8.31%. In terms of profit gains these amount to 91.283 thousand Euros, from which 39.49 thousand Euros are obtained

Table 10.15: Maximum Profit Short Run Peers of Branch B8

			0.45	0.37	0.18
		B8	B2	B33	B28
Inputs	staff	4	3	4	3
	supplycost	13.2181	9.2776	14.5749	15.0787
Outputs	curracc	2009.8513	1161.7352	4127.1885	3188.5107
	othress	4751.0899	2464.5953	9964.8148	10189.6928
	credb	4986.2581	5776.2692	5697.5938	9857.3787
	credass	248.7056	173.3622	422.5467	2806.207
	Profit	151.634	97.3051	309.5077	344.8654

by technical constrained movements and the remaining (51.79) by allocative movements. Profit gains are still higher from allocative movements than from technical movements. This is an expected result because technical movements do not take into consideration factor prices. The above allocative profit gains and allocative efficiency values, therefore, reveal that very small input and output quantity movements are required to considerably increase profit.

Scale effects included in the allocative efficiency are on average higher in the short run assessment than in the long run assessment. Nevertheless, mix effects (averaging 89.92%) are on average the most important component of the short run allocative efficiency value. The scale effect is on average higher than 1 (1.1085) meaning that movements from technical projections to the maximum profit point implied on average decreasing returns to scale movements of bank branches.

Short Run Peers

We can identify two sets of benchmark units in the short run assessment: those that the inefficient bank branch should emulate to become profit maximising given restrictions imposed on input and output changes, and those that the inefficient unit should emulate to become technical efficient given restrictions imposed on input and output changes. Taking branch B8 for illustrative purposes, the peer branches identified by the maximum profit constrained model are shown in Table 10.15.

Branch B8 has 3 benchmark branches (B2, B33 and B28) in the maximum profit constrained model. Note that contrary to the long run case, in the short run maximum profit model the benchmark units do not need to be overall profit efficient. In fact none of the peer branches above is overall profit efficient. This happens because inputs and outputs are not permitted to change by more than a certain amount and so there is no guarantee

Table 10.16: Technical Short Run Peers of Branch B8

			0.52	0.42	0.04	0.01	0.01
		B8	B2	B33	B60	B10	B17
Inputs	staff	4	3	4	6	5	7
	supplycost	13.2181	9.2776	14.5749	18.2361	21.2787	26.3764
Outputs	curracc	2009.8513	1161.7352	4127.1885	4104.8623	5806.227	5677.5172
	othress	4751.0899	2464.5953	9964.8148	8043.4603	14083.404	29726.2398
	credb	4986.2581	5776.2692	5697.5938	19496.8526	12590.4321	17399.0982
	credass	248.7056	173.3622	422.5467	1073.4231	576.7151	514.3504
	Profit	151.634	97.3051	309.5077	369.932	466.3009	663.9061

that the maximum profit plane can be attained.

The peer branches identified by the technical constrained model are shown in Table 10.16.

The short run peers of branch B8 for the constrained technical efficiency assessment are very similar to the constrained maximum profit peers seen before. Branches B2 and B33 play again a major role in defining the peer set of branch B8 meaning that the practices of these bank branches should be seen by branch B8 as role model practices to be emulated both when the constrained technical frontier is to be attained and also when profit is to be maximised.

Note that the peers of branch B8 are all technically efficient branches. This does not need to be so since the technical efficient frontier might not be achievable when input and output targets are restricted from changing by more than a certain amount. This means that non-technically efficient peers might be identified by the constrained technical efficient model. Theoretically this might also happen in the maximum profit constrained model, however as this model is less restrictive (inputs can also increase and outputs decrease) the peer units identified tend to be located on the technological frontier.

10.5 Technical Profit Efficiency Results Over Time

The assessments on profit efficiency involved the computation of overall profit efficiency and its decomposition into technical and allocative efficiency. In order to analyse profit efficiency results over time we decided to focus on technical profit efficiency rather than on overall profit efficiency. This means that no price information is used, and just quantity variables, as shown in Table 10.11, are used to assess technical profit efficiency. No constraints were imposed at this stage on input and output changes, and therefore a long run model was assumed in the analysis of technical profit efficiency over time.

GDF based Malmquist indexes detailed in Chapter 9 are here applied to our sample of bank branches from March 2001 to September 2002. Technical profit efficiency estimates need to be calculated for each branch both in relation to their own time frontier and in relation to another time period frontier. The model used to obtain all efficiency estimates is model (9.15) shown in section 9.5 of Chapter 9.

The technical profit efficiency results obtained independently for each month are shown in Table F.4 in Appendix F where branches are sorted by their average over time efficiency. The results in Table F.4 reveal those units that were consistently efficient over the period of analysis, which was the case of branches B16, B17, B53 and B60.

The GDF Malmquist index analysis uses a base period approach where March 2001 was set as the base period. The reasons for using a base period analysis were already pointed out earlier in this Chapter. Table 10.17 shows the average values for each of the total factor productivity (TFP) change components calculated as detailed in Chapter 9.

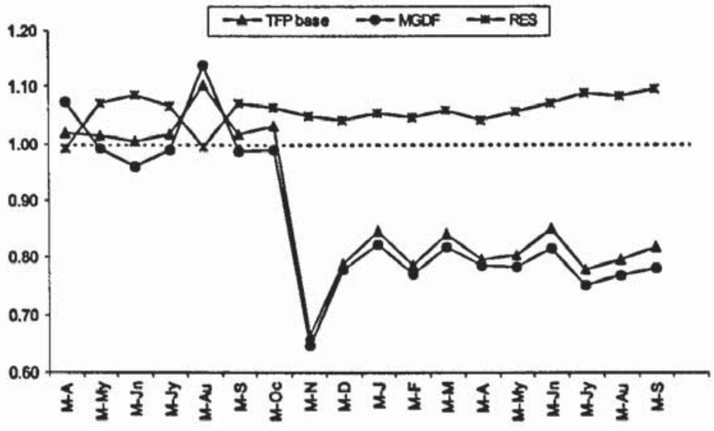
Table 10.17: Base Period Malmquist GDF Index Results

Period	GDF_b	GDF_t	$MGDF_V$	EFCH	THCH	ICH	OCH	RES	TFP
Mar01-Apr01	0.695	0.738	1.0741	1.1415	0.9882	0.9822	1.0152	0.9915	1.0193
Mar01-May01	0.695	0.7063	0.9916	1.0398	0.9682	0.9705	1.0078	1.0703	1.013
Mar01-Jun01	0.695	0.7852	0.9598	1.2209	0.8403	0.9489	0.8947	1.0867	1.005
Mar01-Jul01	0.7037	0.7894	0.9891	1.1963	0.8797	0.9646	0.9174	1.0669	1.0176
Mar01-Aug01	0.7037	0.7823	1.1369	1.1696	1.0132	0.9833	1.0384	0.9944	1.1024
Mar01-Sep01	0.7037	0.7338	0.9865	1.0641	0.9343	0.9531	0.9929	1.0723	1.0176
Mar01-Oct01	0.7929	0.8223	0.9897	1.068	0.9391	0.9566	0.9954	1.0637	1.0322
Mar01-Nov01	0.7929	0.7599	0.6481	0.9854	0.6653	0.6397	1.0502	1.0493	0.6642
Mar01-Dec01	0.7929	0.7443	0.7777	0.9592	0.8193	0.7228	1.1514	1.0418	0.7891
Mar01-Jan02	0.7929	0.7644	0.8219	0.9847	0.8493	0.7447	1.1567	1.0526	0.8449
Mar01-Feb02	0.7929	0.7877	0.7701	1.0144	0.7705	0.6806	1.1449	1.0457	0.7852
Mar01-Mar02	0.7929	0.7938	0.817	1.0257	0.8142	0.7198	1.1485	1.0594	0.8413
Mar01-Apr02	0.7929	0.7716	0.7863	1.0017	0.801	0.6845	1.1872	1.0417	0.7958
Mar01-May02	0.7929	0.781	0.7823	1.0117	0.7835	0.6793	1.1722	1.0564	0.8023
Mar01-Jun02	0.7929	0.7877	0.8145	1.0269	0.8076	0.7197	1.1343	1.0701	0.8497
Mar01-Jul02	0.7929	0.8062	0.7509	1.0506	0.723	0.6686	1.0972	1.0887	0.7791
Mar01-Aug02	0.7929	0.793	0.7694	1.0379	0.7449	0.6816	1.103	1.083	0.7956
Mar01-Sep02	0.7929	0.7933	0.7814	1.0347	0.761	0.7113	1.0891	1.0961	0.8194

We show in Table 10.17 the average efficiency scores obtained for each month (GDF_t) and also for the base period (GDF_b). The average efficiency score for the base period is not the same in all rows of Table 10.17 because some branches closed during the period of analysis and the sample size progressively reduced.

TFP change is mostly around 1 before November 2001 when there was a big decrease in TFP when compared with March 2001. After November 2001 TFP change levels are well below unity implying that input and output levels of March 2001 were better than those of all months of 2002. In Figure 10.5, we show the evolution on TFP during the period of analysis and also the evolution of the two major components of TFP (the Malmquist GDF based index (MGDF) and the residual component (RES)).

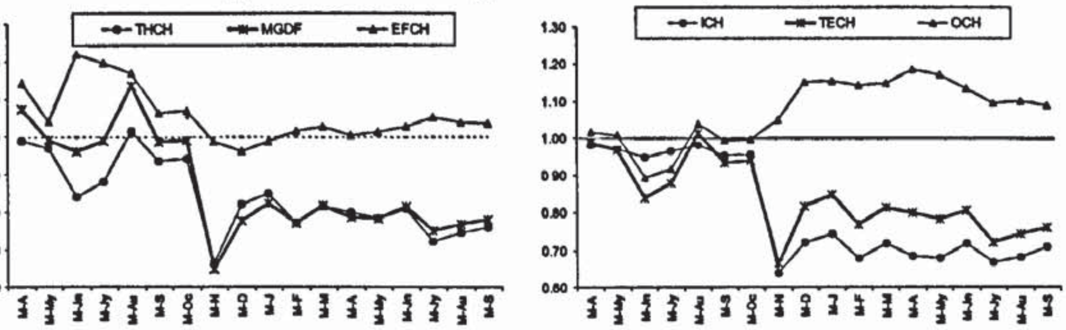
Figure 10.5: Total Factor Productivity Measures



In Figure 10.5 it is clear the big decline in TFP that happened in November 2001 when compared to the TFP levels of March 2001. The main reason behind this decline seems to be supply costs that on average doubled after November 2001, as can be seen in the descriptive statistics presented in Table F.1 of Appendix F. The branch manager was consulted again on the data and told us that values were correct. She did not, however, enter into details concerning the reasons why this happened.

In Figure 10.5 it is also clear that the residual component contributes little to explaining the TFP change. In fact RES values are mostly around 1, and it is the MGDF component that to a greater extent explains the variation in TFP. The MGDF component in turn consists of two components: technological change (THCH) and the efficiency change (EFCH). Efficiency change suffers big fluctuations before November 2001, but after that efficiency change fluctuates little and is very close to 1. In Figure 10.6 this behaviour of the efficiency change component is clear. On the graph at the left we can see that it is technological change that most explains changes in the MGDF, especially after November 2001. Technological change values are almost always below one, meaning that mainly technological regress happened from March 2001 to September 2002, being the exception the month of August 2001. On the graph at the right in Figure 10.6 we can see that technological regress

Figure 10.6: Technological and Efficiency Change



is mostly explained by input deterioration (increase) rather than by output regress. In fact output change only falls well below 1 in June and July 2001, showing a marked increasing trend between October 2001 and May 2002.

Ignoring monthly fluctuations we can say that in the period between March 2001 and September 2002 there was a marked technological regress measured by a value of 0.761 in Table 10.17. TFP change also decreased from March 2001 to September 2002 (see value of 0.8194 in Table 10.17) but at a lower rate than technological change. This is partly explained by positive values of efficiency change in the same period.

Note that in the profit efficiency assessment issues of seasonality were not raised because we found no evidence of the existence of seasonal effects in explaining frontier movements.

10.6 Summary

In this Chapter we present some results regarding the three dimensions of performance introduced in earlier Chapters. Transactional efficiency, operational efficiency, and profit efficiency results were produced for every branch for a number of months (January to September 2002 for the transactional efficiency assessment, and March 2001 to September 2002 for the operational and profit efficiency assessments). Detailed results (efficiency scores, target levels, and efficient peers) from a single month are shown for illustrative purposes, but similar results could be obtained for all other months. In addition, we also analyse efficiency evolution over time since the bank is not particularly interested in knowing whether a bank branch was efficient in a particular month, but rather whether a particular bank branch experienced a consistent behaviour (efficient or inefficient) over the whole period of analysis. As changes in efficiency from one month to another may be explained not only by changes in the relative position of a branch against the efficient frontier, but also by movements of this frontier, it was important to ascertain the degree of technological movements throughout the period of analysis. This was done using Malmquist based indexes that were adapted to the specificities of each assessment.

Conclusions from our time analysis point out for an increase in total factor productivity from January 2002 to September 2002 in the transactional efficiency assessment. Efficiency change is not marked in this assessment and mainly technological change explains the growth in TFP. The best month in terms of technological progress was July 2002. In the operational efficiency assessment there was also a marked technological progress from March 2001 to September 2002, and the months of July 2001 and July 2002 were also the months where biggest progress happened. Technological progress in this case seems, however, to be mostly due to seasonal effects that pushed the frontier upwards. Note that during the Summer there is a big flow of emigrants entering Portugal, and that they only have one or two months during the whole year to to business with their bank in Portugal. This potentially moves the operational efficient frontier upwards during these months and this movement may be confounded with technological change.

The operational and transactional efficiency assessments, though having a different time length, show both technological progress and deem the month of July as being potentially better than the others in transactional and operational terms. Contrary, the profit efficiency assessment showed a high decrease in total factor productivity from March 2001 to September 2002. This seems to be mainly related with an abnormal change in inputs in the month of November 2001. In fact the output change component of technological change shows progress after November 2001. Before November 2001 the best month in terms of TFP growth was August 2001, but after November the best months are January and June 2002.

Chapter 11

Empirical Analysis - Cross analysing Efficiency Results

In the previous Chapter we presented the main results from the transactional, operational, and profit efficiency assessments. In this Chapter we do an integrated analysis of bank branches in all three dimensions. Furthermore, we try to understand and justify some efficiency behaviours of bank branches in light of data that is available but was not directly used in any of the efficiency assessments.

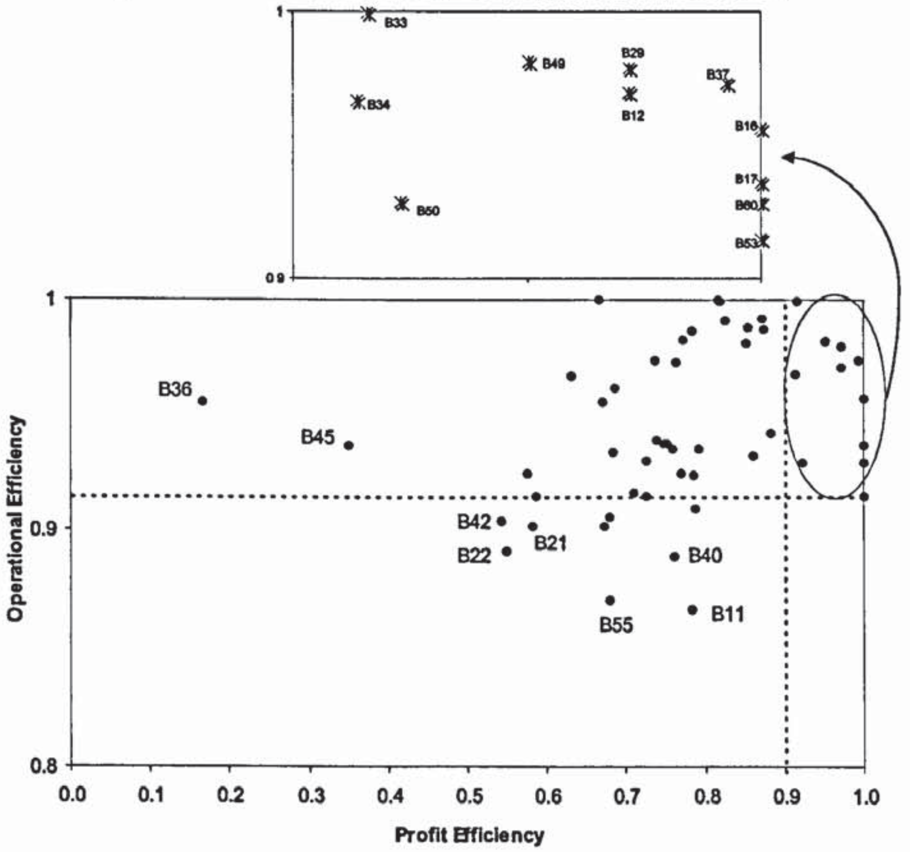
11.1 An Integrated Assessment of Branches Efficiency

Transactional efficiency was computed for the months of January to September 2002. On the other hand operational and profit efficiency were measured for a longer period (from March 2001 to September 2002). This fact implies the loss of some results in terms of the operational and profit assessments when these are to be compared with transactional efficiency results. For this reason, in this cross-efficiency analysis we first focus on the operational and profit efficiency dimensions of bank branches. Later we refer to transactional efficiency in comparison to operational and profit efficiency results. Note that operational and profit efficiency are the most important dimensions of the overall efficiency of bank branches. Indeed, the importance of calculating transactional efficiency is ephemeral because as soon as new distribution channels become effective means of distributing financial services, measuring the efficiency of bank branches in moving transactions to these alternative channels is no longer an issue. Nevertheless, in the short run this dimension of performance is important and is likely to affect other performance dimensions, especially relating to operational practices.

11.1.1 Operational Vs profit Efficiency

Operational and profit efficiency are intended to capture different dimensions of bank branches' activities. In this sense it is possible that some bank branches are good in terms of generating profit but not so good in terms of attracting and maintaining a client base, or vice versa. In order to investigate how bank branches perform in each of these dimensions, results from operational efficiency and profit efficiency assessments are compared. In this comparison we consider average operational and profit efficiency over the period that runs from March 2001 to September 2002 for those bank branches which remained opened during this period (branches that closed at some stage of the analysis are not considered). In Figure 11.1 we show a matrix with both these dimensions. We choose in Figure 11.1 a threshold

Figure 11.1: Profit Efficiency and Operational Efficiency



of about 90% for good efficiency and consider that below this value branches have scope to improve performance. This threshold is arbitrary since the managerial implications of drawing such a graph do not really depend on the chosen threshold, but on the number of units close to the ideal performance (1, 1). Bank Branches with good performance both in profit and operational terms can be classified as 'stars' and they represent benchmarks to be emulated by inefficient branches. Problematic branches are those that have low operational and profit technical efficiency. Special attention should be given to these branches and action is needed to diagnose their problems and to improve their performance.

Bank branches with good profit efficiency and low operational efficiency do not exist in our data set. The absence of bank branches in this quadrant is in line with common sense, since branches with low operating efficiency (i.e. showing a poor performance in selling and attracting new customers) are not likely to be efficient in generating profits. Our results therefore are not surprising.

There are several branches with low profit efficiency but high operational efficiency. If high operating efficiency is maintained by these branches sooner or later these shall show also higher profit efficiency. In fact operational efficiency is measuring the extent to which, given staff and location, branches increase their customer base and their sales while serving their clients. If a branch is good in this task it is likely to have a high value of the various banking products at the bank. That is, outputs on the profit efficiency assessment are likely to reach a reasonable volume, which is good in terms of revenues and profit. Note, however, that in order to be good in terms of profit efficiency inputs (staff and supply costs), which are a source of costs also play an important role.

The correlation coefficient between average profit and average operational efficiency is 0.3. This is not a very high correlation coefficient but it is statistically significant at the 5% level. This means that, in a limited way, higher operational efficiency tends to be related with higher profit efficiency as is noticeable in Figure 11.1.

We present in Table 11.1 the average performance levels of branches found in the star quadrant and those within the 'high operational-low profit efficiency' (HOLP) quadrant both in terms of profit and operational variables. (The detailed values of units in these quadrants can be found in Tables G.1, G.2, G.3, and G.4 in Appendix G). The values in Table 11.1 are averages for the period going from March 2001 to September 2002. Comparing the values of variables for the star units with the same values for the units in the HOLP quadrant we can note that the main differences are on the variables used in the profit efficiency assessment, where star branches using slightly higher levels of inputs produce much more outputs than HOLP branches. In terms of the variables used in the operational efficiency assessment, star branches are better on some variables while HOLP branches are

Table 11.1: Average Characteristics of bank Branches in Two Quadrants

	Profit Efficiency								
	Inputs		Outputs						
	staff	supplycost	depacc	othress	credb	credass			
Star	5.86	35.33	5819.53	17954.14	12808.86	1661.65			
HOLP	5.303	34.87	4463.81	11163.33	9622.41	948.31			
	Operational Efficiency								
	Inputs		Outputs						
	Rent	Staff	ΔCli	Δdepacc	ΔOthre	ΔTdep	ΔCredb	ΔCredas	Trans
Star	2.674	5.862	-37.25	51.38	84.92	-84.34	73.72	39.81	5493.51
HOLP	3.083	5.29	-21.85	29.72	67.05	-66.85	69.96	15.90	4750.10

better on others. Therefore, in terms of growth, branches in each of the star and HOLP quadrants are comparable, but star branches have on average a higher amount of the various types of products. Look for example at branch B19 and compare it with the star bank branch B50 (see details in Table 11.2). In terms of the profit efficiency assessment, B19

Table 11.2: Average Characteristics of Two Bank Branches

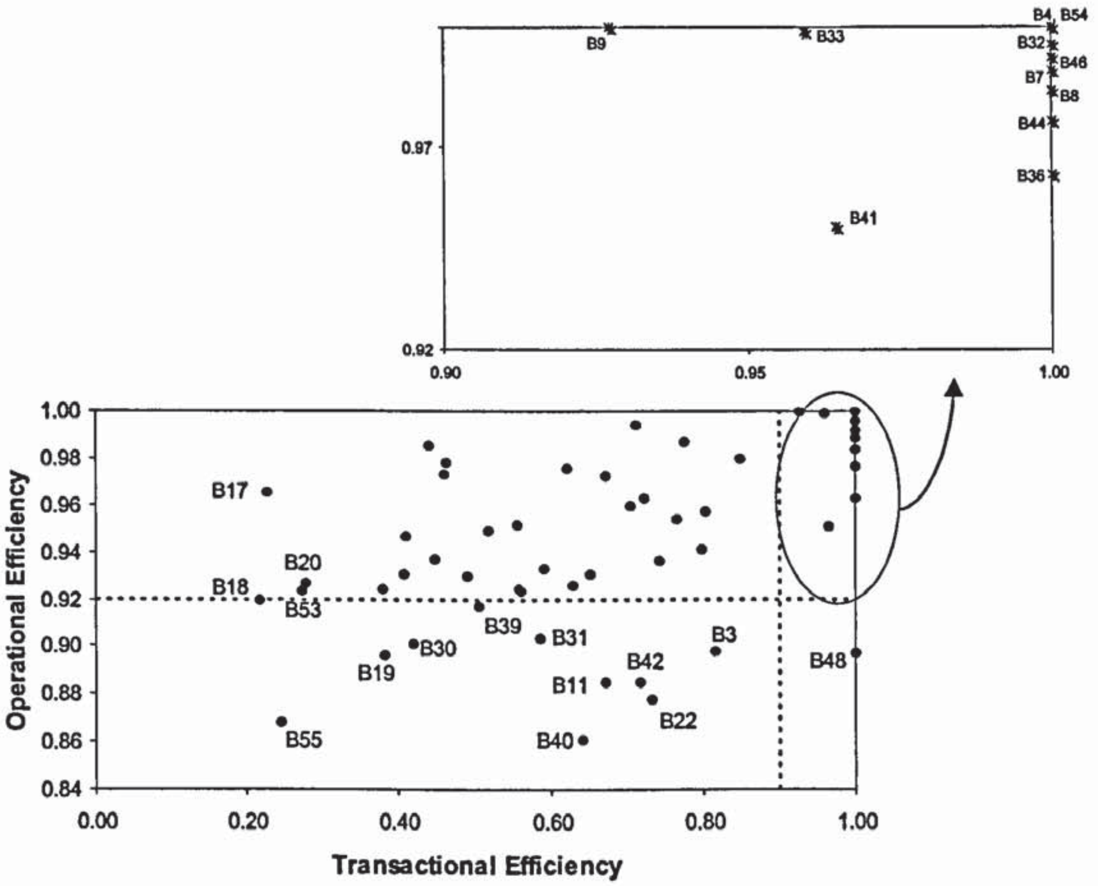
	Profit Efficiency								
	Inputs		Outputs						
	staff	supplycost	depacc	othress	credb	credass			
B50	5	30.298	5171.234	16231.155	9686.86	955.648			
B19	5	35.87	3516.093	12338.948	10434.321	871.681			
	Operational Efficiency								
	Inputs		Outputs						
	Rent	Staff	ΔCli	Δdepacc	ΔOthre	ΔTdep	ΔCredb	ΔCredas	Trans
B50	2.432	5	-20.789	38.764	67.832	-138.924	32.085	3.5	4349.868
B19	4.681	5	-40.722	6.464	-47.838	-86.054	39.865	8	4649.342

has the same number of staff and higher supply costs than branch B50 but produces less of most outputs. In terms of average monthly growth that is measured through operational efficiency both these bank branches are comparable, but in fact branch B19 would need to show higher growth levels if it was to attain the same profit efficiency as branch B50. The profit and operational efficiency assessments are, therefore, complementary since one analyses the growth that bank branches experience from one month to the other, while the other analyses how this growth translates in volume of products that generate revenue to the bank branch. That is, the operational efficiency assessment takes a more dynamic perspective of the activities of the bank branches, while the profit assessment is more static because it looks at stocks rather than at flows.

11.1.2 Transactional Vs Operational Vs Profit Efficiency

The comparison between transactional, operational, and profit efficiency is done for the period from January to September 2002. The relationship that most interests us is that between operational efficiency and transactional efficiency, since profit efficiency is not as related with the ability of bank branches to improve the use of other means of distribution as it is operational efficiency. Using the average efficiency values for this period we present in Figure 11.2 a matrix where operational efficiency is cross compared with transactional efficiency.

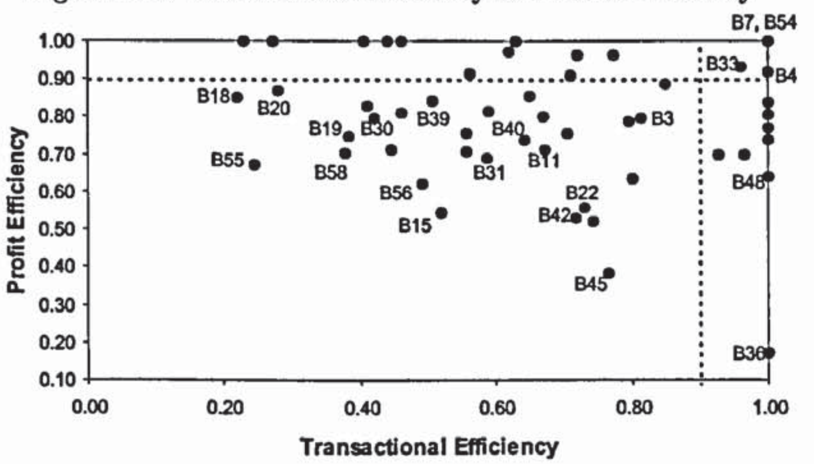
Figure 11.2: Transactional Efficiency and Operational Efficiency



In the same way as before we can see some bank branches in the star quadrant showing high operational and transactional efficiency. There are two bank branches, B4 and B54, that present 100% average efficiency for both dimensions being therefore benchmark branches as far as these two dimensions are considered. There are virtually no branches in the low operational-high transactional quadrant in Figure 11.2, meaning that branches that are good performers in moving transactions to other means of distribution are, in general, also good performers in operational activities that are not transactions related. Indeed the correlation coefficient between these two performance dimensions is 0.46 and this value is statistically significant at the 1% level even if relatively low. This fact confirms our initial hypothesis that moving transactions to alternative means of distributions gives branch staff more time to dedicate to value added activities that lead to increasing sales and the customer base of the bank branch. Since operational efficiency measures the extent to which the branch is able to perform well these value-added activities, the fact that no bank branches can be found in the low operational-high transactional quadrant means that there is a positive relationship between the ability of bank branches to move transactions away from the bank branch and its ability to increase value-added activities.

As far as the relationship between transactional efficiency and profit efficiency is concerned results are as shown in Figure 11.3.

Figure 11.3: Transactional Efficiency and Profit Efficiency



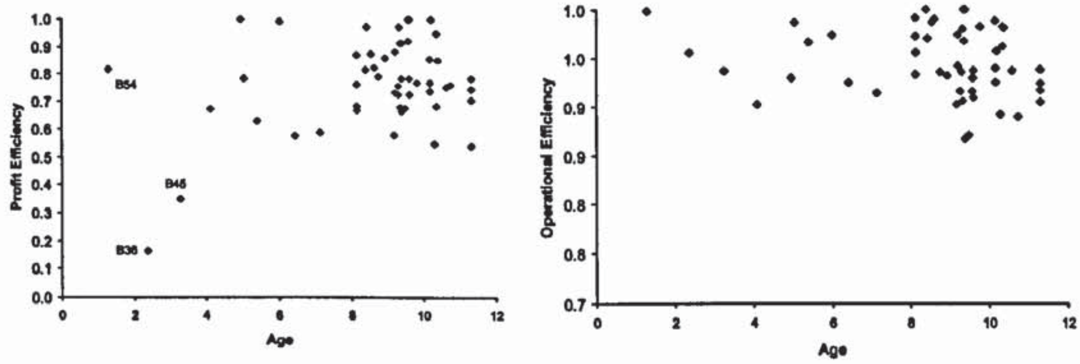
The comparison between profit efficiency and transactional efficiency does not show any significant relationship between the two measures of performance. The correlation coefficient is -0.1848 and this value is not statistically significant. Note that the negative sign is particularly due to two bank branches (B36 and B45) that present very low profit efficiency. This is due to their age (as will be seen in the next section) since they are very young bank branches. Excluding these two bank branches the correlation coefficient is -0.077 and still non-statistically significant. Although there is no correlation between profit and transactional efficiency, we can still identify branches in each of the four quadrants but we can find very different behaviours in terms of each efficiency dimension. For example, for a given level of transactional efficiency we can see branches exhibiting very high or very low profit efficiency levels. On the other hand, there are only 4 branches in the 'star' quadrant, much less than the number found in the previous cross-comparison matrixes. Branch B54 is the only branch showing 100% average efficiency in all dimensions (transactional, operational and profit) for the period between January and September 2002. Apart from this branch, branches B4, B7, and B33 are the only ones presenting high performance levels in all dimensions under analysis, being the star branches when all three dimensions of efficiency are cross-compared.

In summary we can state that branches B4, B7, B54, and B33 can be regarded as benchmark branches from the perspective of all three dimensions of performance, while branches B39, B31, B30, B3, B19, B11, B42, B22, B55, and B40 can be regarded as poor performers in the three performance dimensions.

11.2 Efficiency Vs Age and Competition

It was noted earlier that some environmental variables were not included in the efficiency assessments, though it is possible they affect branch performance. We now investigate how these variables relate to the various efficiency scores calculated. In Figure 11.4 we plotted the profit and operational efficiency dimensions against the age of bank branches.

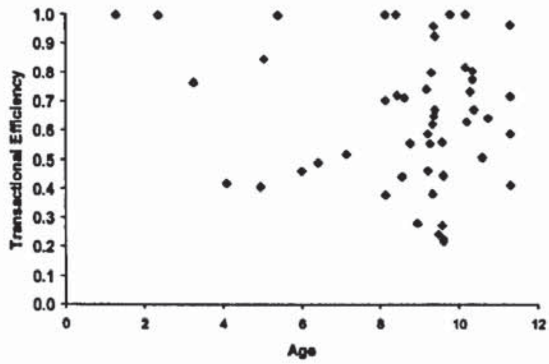
Figure 11.4: Profit and Operational Efficiency Vs Age



If the hypothesis that younger bank branches experience higher growth was true, then one would expect younger bank branches to be more operationally efficient than older bank branches. In fact, as the operational efficiency assessment does not include age, results would be biased towards younger bank branches if age was in fact important. In the right side graph in Figure 11.4 we cannot see any bias of younger bank branches being more operationally efficient and therefore it seems that the exclusion of the factor age did not distort our results (note also that there are not many young bank branches in our sample). The correlation coefficient between age and operational efficiency is -0.2045 but this value is not statistically significant. As far as profit efficiency is concerned, this measures the volume or stock of the various accounts rather than its growth. This means that younger bank branches would in principle have lower profit efficiency because they did not reach a given output volume yet. Three bank branches seem to have been affected by their age in the profit assessment: B36, B45 and B54 (see left side graph in Figure 11.4). These branches are very young (less than 4 years at the beginning of March 2001) and therefore they could not have as much volume in the various accounts as more mature bank branches. Note that these bank branches were included in the set of 'high operational low profit' units discussed in the previous section. The volume of the profit efficiency inputs and outputs of these bank branches (see Table G.3 in Appendix G) is in fact low when compared to other bank branches. This fact is justified by the age of these bank branches and therefore their low profit efficiency shall not ne interpreted as problematic since it is just the result of the evolutionary stage these branches are in. Note that the correlation coefficient between profit efficiency and age is now positive (0.317) and it is statistically significant at the 5% level.

Concerning the relationship between age and transactional efficiency we do not expect this to be marked since age does not seem to be a factor affecting the ability of bank branches to move transactions to alternative means of distribution. This is confirmed in Figure 11.5 where bank branches present high (or low) transactional efficiency irrespective of their age. The correlation coefficient between transactional efficiency and age is negative

Figure 11.5: Transactional Efficiency Vs Age

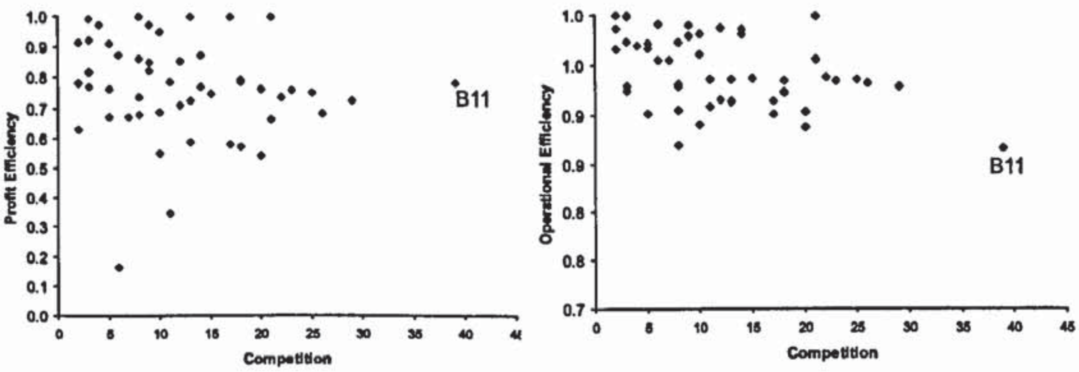


(-0.145) but it is not statistically significant.

In summary age seems to be affecting mainly profit efficiency measures. The relationship between age and profit efficiency is positive, while for the other performance dimensions (operational and transactions) this relationship is negative (although non significant). Recall, however, that our sample has very few young bank branches (only four branches younger than 4 years old), meaning that most branches were old enough for any transitory effects due to the newness to have a significant impact on their efficiency.

As far as the relationship between operational and profit efficiency and competition is concerned we can see in Figure 11.6 that no clear pattern emerges in terms of the effects of competition on the two efficiency measures being considered. We cannot clearly state neither that more competition seems to lead to higher efficiency, nor that low competition enhances efficiency. Note, however, that for one bank branch in particular, B11, the high

Figure 11.6: Profit and Operational Efficiency Vs Competition



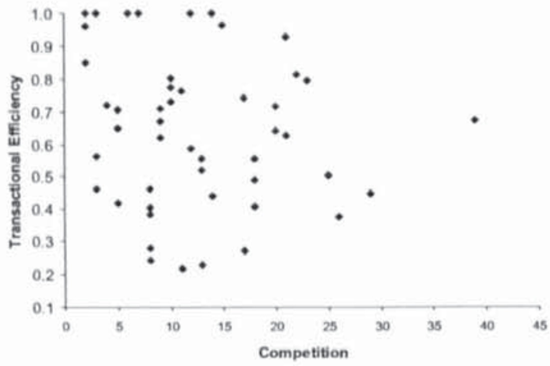
level of competition it faces might be one of the reasons for this bank branch exhibiting low levels of profit and operational efficiency.

Correlation coefficients are negative in both cases and for the case of operational efficiency this is statistically significant (the correlation coefficient is -0.43157). This seems to suggest that the capacity of a branch to attract and maintain clients and increase sales depends in a certain extent on the level of competition it faces.

As far as the relationship between transactional efficiency and competition is concerned we do not expect a high relationship between these two variables. This fact is confirmed in Figure 11.7, where no clear pattern emerges. The correlation coefficient is also negative in this case (-0.22617) although it is not statistically significant.

Note that the above analysis examines the effects of age and competition on the various efficiency dimensions independently. Obviously this type of analysis can fail to identify joint effects of age and competition on the various efficiency dimensions. An analysis of

Figure 11.7: Transactional Efficiency Vs Competition

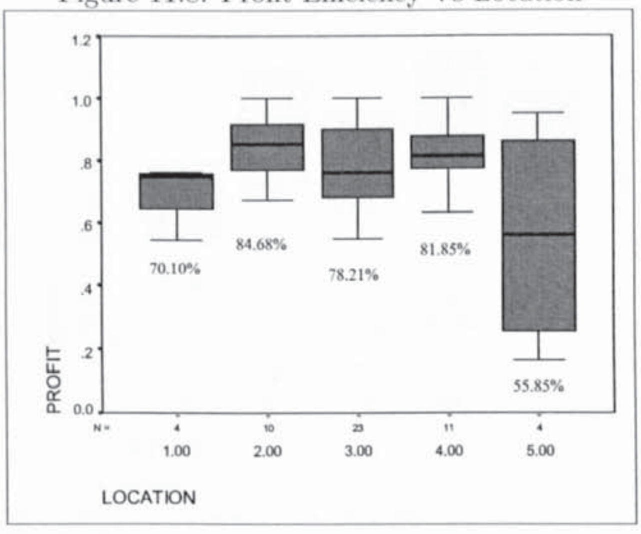


joint effects is postponed to the end of this Chapter.

11.3 Efficiency Vs Location

Apart from the contextual variables examined in the previous session we also analysed the relationship between efficiency and the location of the bank branch. Location is divided into the following 5 possibilities: 1 - Porto city; 2 - Porto region; 3 - Mid size cities, 4 - Other; 5 - Shopping center. In Figure 11.8 the box plot of profit efficiency grouped by location is shown[1]. In this box plot we also show the mean values for each group and it

Figure 11.8: Profit Efficiency Vs Location



can also be seen, on the horizontal axis, the sample size of each group (N). There seems to

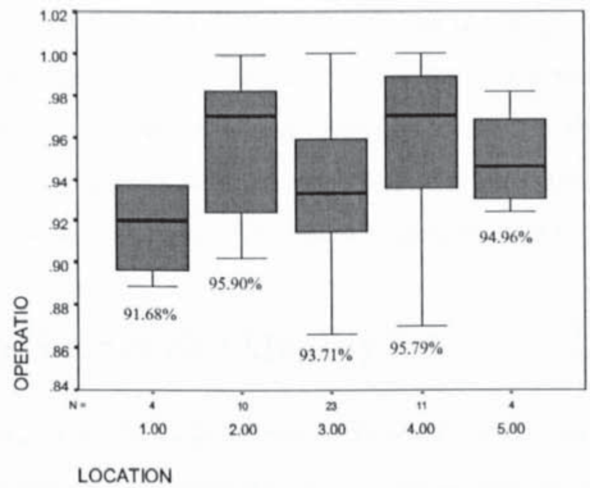
[1]In a box plot the size of the box indicates the distance between the 25% percentile and the 75% percentile of the distribution, with the bar in the middle of the box indicates the median. The lines extending from the box (usually called whiskers) indicate the range of the observations (maximum and minimum values) (Hair JR et al., 1995).

exist differences between locations in terms of profit efficiency, especially as far as location 5 (shopping centers) is concerned. Note however, that there are only 4 observations in this location (the same number of observations happens for location 1), which prevents us from reaching any serious conclusions. Note also that two out of the four branches in location 5, namely those with the lowest profit efficiency, are also two young bank branches (B36 and B45) whose age did not allow yet for a full exploitation of profit potential. This means that we cannot conclude that the profit performance of bank branches in location 5 is poor because half of the units in this location are very young bank branches.

Even though some statistical assumptions are not satisfied by our variable (profit efficiency) we attempted to do some ANOVA tests using SPSS (see Hair JR et al., 1995). Testing for the differences in the 5 locations leads to the conclusion that these are not statistically significant because the F statistic is very small (smaller than F critical). In fact the within groups variance is higher than the variance between groups, meaning that it is unlikely that differences in profit efficiency are due to the location where the branch is in.

In Figure 11.9 we show the box plot when operational efficiency is the variable under analysis. In this case the statistical tests also point to no significant differences in group

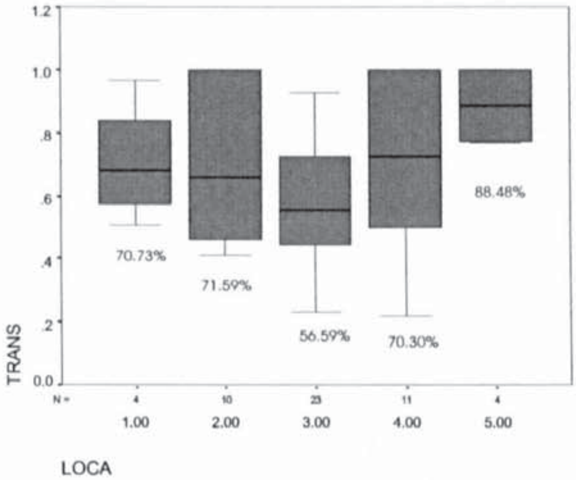
Figure 11.9: Operational Efficiency Vs Location



means, which is confirmed by an inspection of the box plot. Branches in location 1 (Porto city) show the lowest operational efficiency, though we cannot generalise that performance of branches located in this group is worst than that of branches in other locations because there are only 4 observations in group 1. Branches in locations 2 (Porto region) and 4 (other locations) show the highest mean and median operational efficiency, though we cannot conclude for the better performance of branches in these locations given that no statistically significant difference in group means was found in our statistical tests.

In Figure 11.10 we show the box plot when transactional efficiency is the variable under analysis. The location that shows the lowest mean and median transactional efficiency

Figure 11.10: Transactional Efficiency Vs Location



is location 3 (mid dimension cities) and the location that shows the highest mean and median transactional efficiency is location 5 (shopping centers). Though the statistical tests undertaken point to no significant differences in group means, we could easily take as a fact that branches located in a mid dimension city are not in principle as successful as branches located in a big city or its surroundings in moving transactions from the branch to other distribution channels. At the same time branches located in shopping centers might be able to move transactions more easily to other channels, since the profile of clients that use shopping center branches is in principle different from that of typical branch's clients.

11.4 Efficiency Vs Service Quality

In Chapter 5 we explored the link between efficiency, profit and service quality. We argued that an indirect way of incorporating quality into an efficiency analysis was through the calculation of profit efficiency. Here we compare a service quality index with profit, operational, and transactional efficiency measures calculated previously.

It is recalled that the bank measures service quality periodically by means of questionnaires sent to the branch's customers. Given the information collected the bank constructs a monthly service quality index that is then used in a twelve month moving average that shows the pattern on service quality for each bank branch. The service quality indexes we use here are based on the moving average provided by the bank. In Figure 11.11 we show the scatter plot of operational efficiency and service quality. The data on service quality relates to September 2002, meaning that it is an average of SQ indexes from September 2001 to September 2002. The average values calculated for operational and profit efficiency measures also consider this year period.

Figure 11.11: Operational Efficiency Vs Service Quality

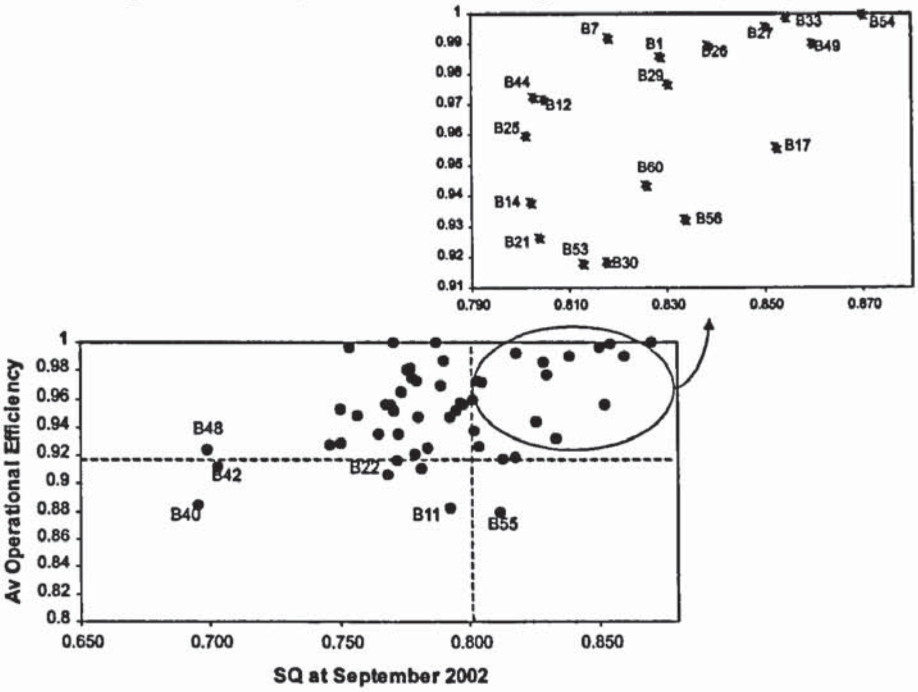
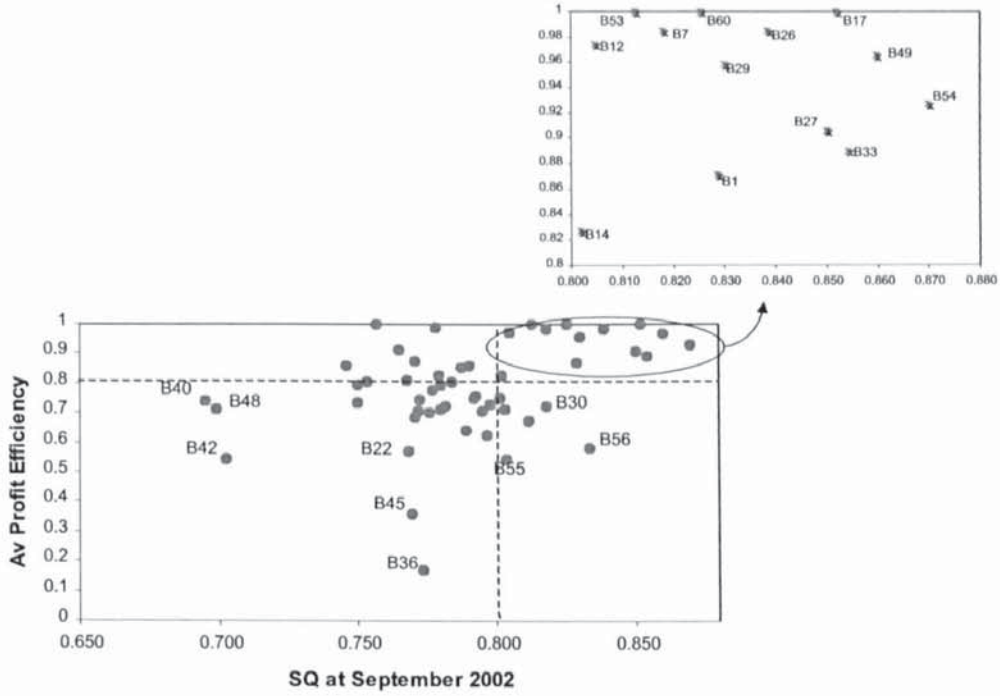


Figure 11.11 shows a link between operational efficiency and service quality which indicates a positive relationship between the two variables. The correlation coefficient is 0.43 and this is significant at the 1% level. More important than analysing this relationship is, however, the analysis of the bank branches that lie on each of the quadrants identified in Figure 11.11. It is interesting to note that there are almost no branches on the quadrant with low operational efficiency and high service quality, which obviously is in accordance with the common sense feeling that bank branches providing high service quality are more effective in increasing their customer base and in selling banking products. There is a set of bank branches that can be seen as benchmarks in operational terms, which are those that show at the same time high operational efficiency and high service quality. Such branches attain growth rates that are better than those of other branches, while at the same time they tend to serve clients with a better service quality than other bank branches.

11.4 Efficiency Vs Service Quality

Comparing the same service quality index with profit efficiency the picture is as shown in Figure 11.12. Here the relationship between service quality and profit efficiency is not

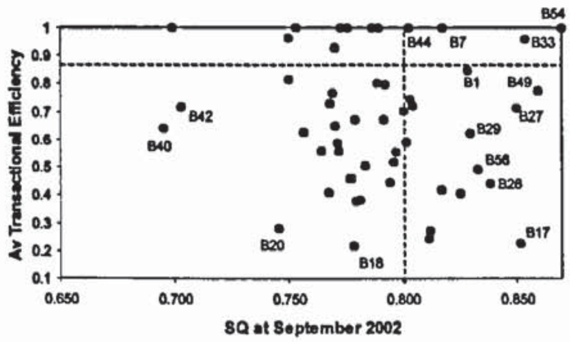
Figure 11.12: Profit Efficiency Vs Service Quality



as clear as before. We can find now some branches in the quadrant with low profit efficiency and high service quality. Nevertheless, there appears to be a positive relationship between profit efficiency and service quality, meaning that higher service bank branches tend to be good performers both in terms of generating profits as in terms of operational growth. In fact the correlation coefficient between service quality and profit efficiency is lower than before (0.371) but this value is still significant at the 1% level. The weaker relationship between profit efficiency and service quality may be due to a lag effect between the costs of quality and the revenues accruing from superior quality. According to Kordupleski et al. (1993) the effects of quality improvements on market share depend on the length of the purchase cycle. In banking it might take a long time to translate quality improvements (or deterioration) into market share and profit improvements (deterioration), since the quality image the customer builds may require several contacts with the bank to be clearly formed in the client's mind. The effects of service quality improvements might, therefore, be more immediate in terms of improving operational efficiency, while profit efficiency improvements are likely to take longer to happen. A detailed analysis of eventual lag effects of service quality improvement on the two performance dimensions is not, however, within the scope of this study.

The relationship between service quality and transactional efficiency is completely different from that shown before, as there appears to be no relationship between these two performance dimensions. This can be seen in Figure 11.13.

Figure 11.13: Transactional Efficiency Vs Service Quality



The correlation coefficient between transactional efficiency and service quality is -0.0914 and this value is not statistically significant. This fact seems to suggest that the efficiency by which bank branches move transactions away from the bank branch does not affect service quality. This is an important finding since it means that customers do not associate the increasing use of alternative distribution channels (and the decreasing use of the branch for transactional purposes) with losses in terms of service quality (nor gains).

11.5 Joint Effects of Contextual Factors on Efficiency

In the previous sections we analysed the relationship between efficiency and a number of contextual factors such as age of the branch, level of competition, location, and service quality. Each of these factors was analysed independently, which may be seen as a caveat of our analysis. For this reason, we decided to analyse the joint effects of contextual factors on the three measures of efficiency. For this purpose we used a Tobit model or censored regression model, which handles cases where the dependent variable lies within a certain range (0 to 1 in this case). Tobit models have been used by a number of authors in the DEA literature in various contexts. For example, Kerstens (1996) explained efficiency distributions of French urban transit companies through a Tobit model, Kirjavainen and Loikkanen (1998) explained the inefficiency of Finnish secondary schools through a Tobit model, and Chilingerian (1995) explained DEA efficiency scores of physicians through a Tobit model.

Tobit models are preferred to ordinary regression analysis to explain DEA efficiency scores, because the dependent variable is censored. Nevertheless, the use of DEA efficiency scores as the dependent variable of Tobit models does not exactly fit the underlying theory of these models (see e.g. Chilingerian, 1995). One of the reasons for this is the fact that DEA efficiency scores are not independent between each other. Simar and Wilson (2003) mention this problem in Tobit analysis of DEA scores, and propose a bootstrapping approach to solve it. As our aim in using the Tobit model is just to get a general feeling about the joint effects of a set of variables on the DEA efficiency scores, we believe a more elaborate analysis as that proposed by Simar and Wilson (2003) is not justified given our general objective.

We used LIMDEP to estimate three Tobit models for each efficiency dimension (transactional, operational and profit). Efficiency scores are usually converted into inefficiency scores because the Tobit model is censored at zero. We did not perform this conversion because LIMDEP allows for the possibility of censoring both tails. We choose this alternative, using a lower limit of zero and an upper limit of one. The independent variables used in the Tobit model were, age, competition, and service quality[2]. In order to check for non-linear relationships between these variables and efficiency we also considered squared terms of the above explanatory variables. Only in some cases these terms were included in the Tobit model as can be seen in Table 11.3, where Tobit results are shown[3].

Table 11.3: Results from Tobit models

	Profit Efficiency		**Operational Efficiency**		**Transactional Efficiency**	
	Coefficient	p-value	Coefficient	p-value	Coefficient	p-value
Constant	8.95	0.1595	0.725	0.0000	2.075	0.0312
SQ	-24.25	0.1363	0.289	0.0217	-1.398	0.2132
Age	0.139	0.0056	0.1914E-04	0.9287	-0.0331	0.0841
Comp	-0.631E-04	0.6704	0.1682E-04	0.5962	0.362E-04	0.8941
SQ^2	16.62	0.1073				
Age^2	-0.0076	0.0366				
$Comp^2$			-0.476E-04	0.0056		
Log-Lik	**15.903**		**99.5091**		**-14.544**	
p-value	**0.0711**		**0.1308**		**0.628**	

According to the Tobit model on the profit efficiency scores, the factor appearing to

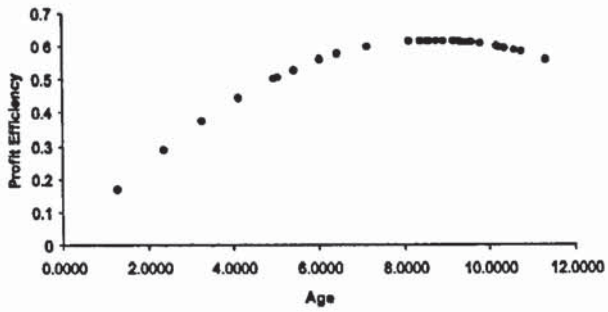
[2]Location was not considered because the inclusion of this variable (in the form of four dummy variables) decreased the overall quality of the three Tobit models. In addition the p-values of the dummies were jointly non-statistical significant for the three models.

[3]The non-linear terms included in the Tobit models in Table 11.3 are only those that showed some level of significance in explaining the dependent variable.

explain better profit efficiency is the age of the bank branch (Age and Age^2). All the remaining variables are not statistically significant at the 5% level, although the service quality, as far as its non-linear component is concerned, is statistically significant at the 10% level.

The Tobit model identifies both a negative (Age^2) and a positive (Age) impact of age on profit efficiency. This means that a non-linear relationship is identified between these two variables. The theoretical relationship modelled by the Tobit model between the age of the branch and profit efficiency is as shown in Figure 11.14. That is, as age increases

Figure 11.14: Theoretical relationship between age and profit efficiency



profit efficiency is likely to increase until a certain point, but after a certain age (about 9 to 10 years old) the profit efficiency is likely to start slowly decreasing. This is an interesting relationship identified by the Tobit model, as it models the life cycle of a bank branch with a growth period, a maturity period and a declining period after a certain age.

As far as service quality is concerned, one would expect a positive impact of this variable on profit efficiency. This happens for the SQ^2, but not for SQ, whose impact on profit efficiency is negative. This also suggests a non-linear relationship between SQ and profit efficiency although in this case the reliability of such a relationship is not supported by statistically significant coefficients.

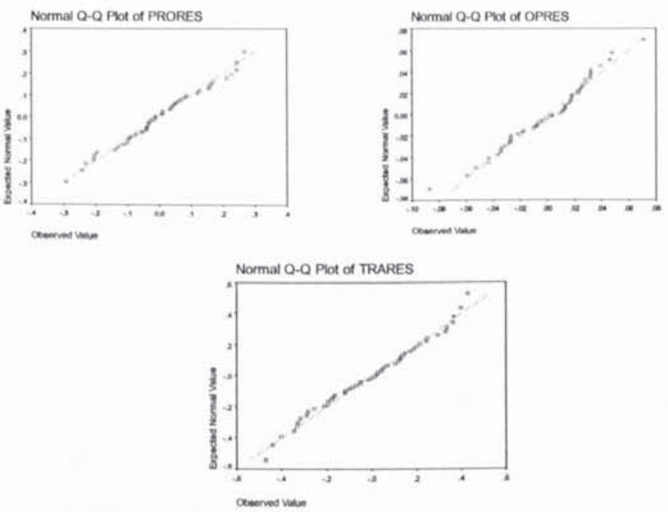
Concerning the Tobit model that explains operational efficiency the most important factor is service quality, which is statistically significant at the 5% level. The relationship is positive as expected. We did not consider in this case the non-linear factor on service quality as it was not statistically significant. Competition seems to affect operational efficiency as far as its squared component is concerned. This effect is negative as one would expect. That is, the higher level of competition seems to be related with lower levels of operational efficiency.

As far as the Tobit model that explains transactional efficiency is concerned, none of the variables seem to have a significant impact on explaining differences in transactional efficiency. This finding is consistent with the previous analysis we have done, when each contextual factor was analysed independently.

Concerning the overall quality of our Tobit models, we calculated the difference between the Log-likelihood factor in Table 11.3 and a Log-likelihood factor obtained when all coefficients of the independent variables are assumed to be zero. This difference follows the χ^2 distribution, and the p-values obtained from this distribution are shown in Table 11.3. The p-values show that overall the quality of our models is not very high, with the profit efficiency model being significant at the 10% level, and the operational efficiency model being significant at the 13% level. For the transactional efficiency model the overall quality is even worse as the p-value is about 0.628.

In order to check for the adherence of some of the assumptions behind the Tobit model to our data set we did some graphical tests. One of these tests was to check for the normality of the residuals. We did some normal probability plots to do this and the results are as shown in Figure 11.15. For the three dependent variables considered (profit efficiency,

Figure 11.15: Normal Probability Plots



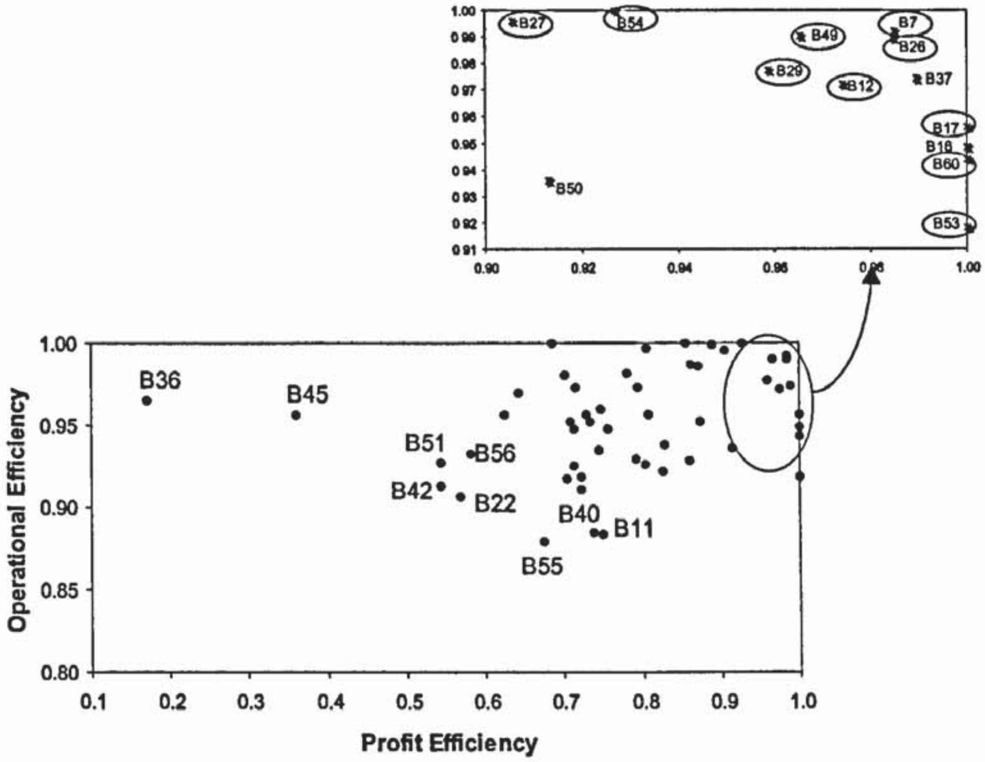
operational efficiency and transactional efficiency) there are some deviations from normality and a number of outliers are identified. However, in none of the cases we can classify these deviations from normality as being much severe. We also did some plots of the residuals against each independent variable to check for unequal variances (heteroscedasticity), but there was no evidence that this existed.

11.6 Results from our Study Vs Pre-Conceptions of the Bank

Having produced the above results we discussed them with the manager of the network of branches assessed in order to find out whether our results accord with perceived wisdom of the bank. What we were interested in was mainly to know the extent to which the bank branches identified as most efficient on this study were in fact so from the perspective of management. For example in calculating transactional efficiency most of the variables that should ideally be used could not be provided and therefore we used other type of variables. This was an important limitation in measuring transactional efficiency and therefore it was important to ascertain whether the branches that we identified as most efficient in transactions were in fact considered so by the bank too. We showed the network manager the results on transactional efficiency (average values from January 2002 to September 2002) and the manager agreed with our classifications for almost all branches. For example it was agreed that branches B17, B18, B55, B20 and B53 were very poor in their role of transferring transactions for other distribution channels, whereas branches B4, B54, B32, B46, and B7 were very good in performing this role. Some of these branches are located in rural areas and the manager told us that despite of this fact some of these bank branches showed a very good performance in their transactional role.

Operational and profit efficiency performances were also showed to the network manager. In this case a certain confusion arose between profit efficiency and profit in absolute terms. When we said that a bank branch had high profit efficiency there was the tendency to associate this with very high actual profits. It was important to clarify concepts here because high profit efficiency need not correspond to high profit. It simply means that given the staff and the supply costs of the bank branch (the two inputs considered in the profit assessment) it is achieving a volume of the various products that is higher than those of branches in similar conditions. Having clarified concepts it was agreed that most of the bank branches that we identified as good performers in terms of profit efficiency, operational efficiency, and service quality were in fact considered best performers by the bank. The data we used for showing best and worst performers was an average of efficiency scores from September 2001 to September 2002. Figure 11.16 considers average operational and profit efficiency values for this period, and identifies for the best performance group those that had high service quality levels with a circle. Our focus of discussion with the network manager was therefore based on the three performance dimensions shown in this figure. It was decided not to include here the transactional dimension because as we saw

Figure 11.16: Operational Efficiency Vs Profit Efficiency



before it is not correlated with profit efficiency and service quality. Therefore considering best performers on the 4 dimensions (profit, operational, transactional and service quality) would result in very few bank branches appearing as good performers overall.

It is worth noting that in analysing performance on the three dimensions mentioned above it was found that the manager agreed more strongly on the worst rather than on the best performers. For example the network manager showed no doubts in considering branches B40, B42, B22, and B55 as bad performers. Some reasons were pointed out for this, since problems at these bank branches were well known to management. For example, one of these bank branches is located in the center of Porto and it was recognised that technology at this branch was obsolete, and that there was a high difficulty in moving clients to use other distribution channels. This means that branch staff are mostly occupied with serving clients on general transactions and there is no much time left for selling. At the same time it was recognised that personnel at this bank branch was not pro-active in trying to sell and capturing new clients. Thus at this bank branch transactions are still the main activity carried out by staff. For another bank branch (B22) in the worst performer group it was said that this branch got several clients from another branch that closed and clients were not satisfied by this change, which was clearly influencing the overall performance of that branch.

As far as best performers are concerned agreement between our results and managerial perceptions was weaker. In some cases branches identified as the best performers were located in small rural cities and management did not expect to see these bank branches appearing as best performers. This mainly arose due to the fact that the business volume at these branches was not very high and therefore this type of bank branch was not seen as contributing much to the profit of the bank as a whole. For example branch B54 shows the highest service quality levels from September 2001 to September 2002 and also very high operational and profit efficiencies for the same period. This bank branch is however small (only 3 staff members) and is located in a small city, meaning that it was not expected to be considered in the best performers group by the network manager. It was however recognised that this bank branch was in fact a good performer given its size and the external conditions it faces. Note that this bank branch was also very good in transactional terms being in fact a best performer in all four dimensions considered.

One of the most important disagreements related to branch B11 which was seen by the network manager as a role model branch. In our case this branch was classified in the group of the worst performers since it exhibited low operational and profit efficiency and also low service quality (note however that the values in these performance dimensions were not the lowest (nor close to the lowest) observed). In terms of transactional efficiency too this bank branch is not amongst the best. The reasons for this disagreement were not completely sorted out, but one of the issues that came out related to the transactional load at this bank branch that was in the opinion of the network manager higher than our values showed. For example, we have values for transactions that show that on average branch B11 performs less transactions than branch B10 and B16. This fact was considered strange by the network manager, and therefore some error might exist in the data we have used. Our variable concerning number of transactions at the bank branch is a sum of the following transactions: cheque deposits, cash deposits, cash withdrawals, payment of bank cheques, and internal transfers. It might happen that other transactions not considered here are used by the bank to assess the transactional volume of the bank branch and therefore we are not considering all the transactional load of bank branches. Under this circumstance bank branches that show a high number of those transactions not considered in our assessments might indeed appear as inefficient when they are not. This might have been the case of branch B11.

In a study like the one that was undertaken here it might happen that in some cases the information concealed from the analysis is more important than the information actually included there. It is impossible to consider all possible variables that explain differences in performance between bank branches. It is therefore important to keep in mind that our results only reflect differences between bank branches on the variables that were considered on the efficiency assessments. This means that many reasons might exist for explaining why some bank branches were considered less efficient than others. Such reasons should be looked at in terms of the variables that were not considered in the assessment, which for some cases might be more important than those considered. Apart from these limitations, which are inherent to any quantitative study, we believe that this study indeed captured important performance differences between bank branches and mostly classified them rightly. In addition, the relationships identified between the various performance dimensions confirmed our initial expectations regarding the link between our efficiency measures.

Finally note that the fact that managers agreed more with the identification of worst performers can be linked with the fact that DEA by nature makes a stronger identification of weak performers. It tries to show the branch in the best light and if it even then finds it inefficient this means the branch is, more often than not, truly inefficient. This does not hold for efficient branches as they may appear so through an odd combination of input-output levels and lack of similar comparators.

11.7 Summary

In this Chapter we linked the three dimensions of performance that have been analysed independently in the previous Chapter. From this linkage we were able to show those bank branches performing particularly well on the operational and profit dimensions, on the operational and transactional dimensions, and also on the profit and transactional dimensions. Comparing pairwise the three efficiency dimensions we could conclude for a positive relationship between operational and profit efficiency and between transactional and operational efficiency. In contrast transactional and profit efficiency do not seem to be related.

In this Chapter we also showed the relationship between the three performance dimensions and some contextual factors not included in the analysis. It was shown that the age of the bank branch affects particularly its profit efficiency in a positive way, the level of competition affects particularly its operational efficiency in a negative way, and the location of a bank branch does not seem to affect its performance at any of the three levels. As far as the relationship between service quality and the three performance dimensions is concerned it was shown that there is a positive relationship between service quality and profit efficiency and between service quality and operational efficiency. There seems to be no relationship between service quality and transactional efficiency, which suggests that the customers do not value negatively, nor positively, the efforts of bank branches in moving transactions away from the branch to other distribution means.

Joint effects on efficiency of contextual factors were also analysed through Tobit models. Results from these models are in close agreement with the analysis performed individually for each contextual factor, with age being the factor that most explains profit efficiency, and service quality and competition being the factors that most influence operational efficiency. As far as transactional efficiency is concerned, none of the contextual factors considered seem to contribute much to explaining differences in transactional efficiency across branches.

Chapter 12

Conclusion and Further Developments

This Chapter concludes this thesis. It is divided into two main sections. In the first section we present the main theoretical contributions of this thesis to the literature, mentioning also the main findings regarding our empirical application to bank branches. In the second section we present directions for future research regarding the measurement of bank branches efficiency.

12.1 Contributions to the Literature

In this thesis we have made both methodological contributions and applied them to a real life example. The methodological contributions concern the development of models to measure efficiency when some inputs and/or outputs are negative, the development of models to measure profit efficiency, and the development of models that provide targets that are nearer to the production unit being assessed than targets yielded by traditional models.

All methodological contributions in this thesis were guided by our empirical application to bank branches, whose changing role from transactions based to sales based led to innovative ways of assessing their efficiency. At the same time, it was also important to reflect in our efficiency assessments the objectives of bank branches in the perspective of the bank managers, and also to account for the fact that bank branches are service and for-profit organisations. As a result we developed models to measure the efficiency of bank branches in three efficiency areas of interest: transactions, operations, and profit. Transactional efficiency is intended to capture the ability of a bank branch to move transactions from the bank branch to alternative distribution channels. This objective was understood by bank and branch managers of utmost importance given the advantages for the bank as a whole of increasing the usage of alternative distribution channels. Operational efficiency is intended to capture the extent to which a bank branch accomplishes its operational objectives of attracting and maintaining customers, increasing sales, and serving clients. Finally, profit efficiency is intended to capture the extent to which a bank branch is good at generating revenues given the costs it faces. Each of the above efficiency measures contributes to explaining a part of the overall activity of bank branches, therefore complementing each other to provide a wider picture of the performance of bank branches. To the authors knowledge, this is the first study that considers explicitly a transactional efficiency measure, and complements it with the operational and profit dimensions of bank branches.

In order to measure the above efficiency measures some theoretical developments were needed. Theoretical developments needed for measuring operational efficiency relate with DEA models to deal with negative data. We propose in Chapter 6 a new model (Range Directional Model - RDM) for dealing with this type of data that has some advantages over existing approaches. Firstly, the proposed approach results in an efficiency score that is radial with respect to an ideal point (and not with respect to the origin as customary), and secondly the proposed approach allows for two different routes to be followed towards the efficient frontier depending on the objectives of the unit concerned. One route gives priority to changing factors on which the production unit performs best, and the other gives priority to changing factors on which the production unit performs worst. Given that the RDM model results in radial efficiency scores this opened up the possibility for calculating Malmquist indexes when data are negative. To the authors knowledge this is the first attempt to use Malmquist indexes in situations where negative data exist.

Theoretical developments for measuring profit efficiency were also required since the literature using DEA to assess this type of efficiency is still limited. We propose in Chapter 8 a new efficiency measure: Geometric Distance Function (GDF), that can be used to calculate overall profit efficiency and to decompose it into its technical and allocative components. The GDF approach is developed in that Chapter both under the long run and short run assumptions. The advantage of the GDF over existing approaches is the fact that it is a measure capable of incorporating all the sources of inefficiency and it is also a decomposable measure. In addition, it measures profit efficiency based on the adjustments required on the inputs and outputs to move from an actual point to a maximum profit point, therefore avoiding the use of ratios between profit levels that could render negative efficiency measures in case profit was negative.

In Chapter 9 we further extend the use of the GDF to calculate total factor productivity (TFP) change and to decompose it into its efficiency change, technological change, and a residual component that is scale related. The main differentiating factor regarding our approach is that the GDF is adapted to calculate (TFP) based just on observed data. This means that no distance functions are used as a means to calculate TFP in the multiple input/multiple output case.

Another development put forward in this thesis (Chapter 7) is that of calculating closest targets when non-oriented measures are used in efficiency assessments. This development allows one to keep the original spirit of DEA of showing each production unit in the best possible light in a non-oriented framework. Indeed, most of the models used to measure efficiency in a non-oriented space look for targets that the the furthest to the assessed unit rather than the closest. This is not in line with intuition nor with the way management exercises judgment in general. The framework developed in Chapter 7 is applied in Chapter 8 to calculate technical profit efficiency.

The application to a sample of bank branches of the three efficiency measures proposed in this thesis reveals some interesting insights into the functioning of bank branches, especially when these dimensions are cross compared. For example, our results suggest a positive link between operational and profit efficiency, and show a number of bank branches that can be considered excellent in terms of their operational and profit objectives. At the same time we found that transactional efficiency is positively related with operational efficiency, but no significant relationship arises between transactional efficiency and profit efficiency. This is a result consistent with expectations since moving general transactions from the branch to other channels (measured by transactional efficiency) is likely to affect especially the growth of various accounts and clients at the bank branch (measured by operational efficiency) since branch's staff have more time to dedicate to these activities when transactions are performed through other means. We also analysed all performance dimensions in relation to service quality, and found positive links between service quality, operational efficiency, and profit efficiency. On the contrary, service quality seems to be unrelated with transactional efficiency, meaning that customers do not value negatively the efforts of the bank branch in moving basic transactions away from the branch.

Our empirical results were presented to the manager of the network of branches under analysis and in general our assessments resulted in classifications that accorded prior views on the performance of bank branches. This happened irrespective of the several data limitations that were pointed out throughout this study, which reveals the adequacy of the models used to assess bank branches' performance.

12.2 Further Developments

Further developments to this thesis are possible in a number of respects. Concerning the theoretical developments introduced in this thesis further work can be done on the RDM model to deal with negative data, on the GDF model to measure profit efficiency, and on the closest targets (CT) procedure.

Under the RDM model we can obtain different targets depending on the directional vector chosen. One of the directions that we specify tends to generate targets that are close to the closest targets that could be found. This, shows that depending on the directional vector chosen we can in a way manipulate the closeness of the resulting targets. Therefore, further developments on this issue would be the manipulation of the directional vector such that RDM models could in fact provide the closest targets (according to a certain criterion) to production units. At the same time the introduction of preference information in the RDM model could also be a means of achieving this objective.

The RDM model is also used in this thesis to calculate Malmquist type indexes in the presence of negative data. The resulting approach still suffers from some problems, like the definition of an ideal point based on which the range of possible improvement is to be calculated, and the possibility of infeasible models. Developments on this issue should pass through solving the above problems.

Further work is also possible on the closest targets procedure, especially as far as attempts to simplify the efficiency measures and the procedures for finding all efficient facets of the production frontier are concerned. We used a given criterion of closeness in the CT procedure, but other criteria can also be explored.

The GDF approach developed in this thesis to calculate and decompose a measure of overall profit efficiency can also be further investigated. This should include some further analysis on the economic interpretation of the resulting efficiency measures, since these measures express adjustments in terms of input/output quantities but do not contain information concerning relationships between profits. At the same time the decomposition of allocative efficiency into scale and mix effects can also be the subject of further work. The approach we propose starts by isolating scale effects, and mix effects are then found residually. A reverse order for this decomposition would imply different results that could be worth investigating.

The GDF approach is extended in this thesis to calculate total factor productivity (TFP) change and to decompose it into a number of components. One of these components is called residual component and we show that it is scale related. The interpretation of this component is, however, complicated and we do not fully investigate it in this thesis. Further work is, therefore, in need regarding this subject.

Concerning the empirical application to bank branches some more work could also be done at this level. This would pass through a deeper involvement of the bank and branch managers in order to try to generate some guidelines for branches to follow in order to improve their efficiency. On the one hand a detailed analysis of each assessment should be performed with the bank and branch managers in order to identify data limitations and try to correct for these limitations so that assessments better reflected management views. On the other hand, causes for good and poor performance should be analysed in more detail with bank and branch managers, so that some remedies could be pointed out to those branches that in fact showed some problems. The results obtained from our empirical work provide useful information concerning the performance of bank branches but do not really point out for directions to be followed by bank branches. From a closer involvement of the bank in this empirical analysis such directions could be provided to bank branches.

References

Achabal, D. D., Heineke, J. M., and McIntyre, S. H. (1984). Issues and perspectives on retail productivity. *Journal of Retailing*, 60(3):107–127.

Adams, R. M., Berger, A. N., and Sickles, R. C. (1999). Semiparametric approaches to stochastic panel frontiers with applications in the banking industry. *Journal of Business and Economic Statistics*, 17(3):349–358.

Agrell, P. and Bogetoft, P. (2001). DEA-based regulation of health care systems. Paper presented at the 7th European Workshop on Efficiency and Productivity Analysis, Oviedo, Spain, 25-29 September 2001.

Agrell, P. J., Bogetoft, P., Brock, M., and Tind, J. (2001). Efficiency evaluation with convex pairs. Paper presented at the 7th European Workshop on Efficiency and Productivity Analysis, Oviedo, Spain, 25-29 September.

Al-Faraj, T. N., Alidi, A. S., and Bu-Bshait, K. A. (1993). Evaluation of bank branches by means of data envelopment analysis. *International Journal of Operations and Production Management*, 13(9):45–52.

Ali, A. I. and Seiford, L. M. (1990). Translation invariance in data envelopment analysis. *Operations Research Letters*, 9:403–405.

Ali, A. I. and Seiford, L. M. (1993a). Computational accuracy and infinitesimals in data envelopment analysis. *INFOR*, 31(4):290–297.

Ali, A. I. and Seiford, L. M. (1993b). The mathematical programming approach to efficiency analysis. In Fried, H. O., Lovell, C. A. K., and Schmidt, S. S., editors, *The measurement of productive efficiency: Techniques and Applications*, pages 120–159. Oxford University Press, New York, Oxford.

Allen, K. (1999). Dea in the ecological context - an overview. In Westermann, G., editor, *Data Envelopment Analysis in the Service Sector*, pages 203–235. Gabler Edition Wissenschaft.

Allen, R., Athanassopoulos, A., Dyson, R. G., and Thanassoulis, E. (1997). Weights restrictions and value judgements in data envelopment analysis: Evolution, development and future directions. *Annals of Operations Research*, 73:13–34.

Allen, R. and Thanassoulis, E. (1996). Increasing envelopment in data envelopment analysis. Warwick Business School Research paper N. 216.

Aly, H. Y., Grabowski, R., Pasurka, C., and Rangan, N. (1990). Technical, scale, and allocative efficiencies in U.S. banking: an empirical investigation. *The review of economics and statistics*, 72:211–218.

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

Anderson, E. W., Fornell, C., and Lehmann, D. R. (1994). Customer satisfaction, market share, and profitability: Findings from Sweden. *Journal of Marketing*, 58:53–65.

Appa, G. and Yue, M. (1999). On setting scale efficient targets in DEA. *Journal of the Operational Research Society*, 50(1):60–69.

Asmild, M., Hougaard, J. L., Kronborg, D., and Kvist, H. K. (2003). Measuring inefficiency via potential improvements. *Journal of Productivity Analysis*, 19(1):59–76.

Athanassopoulos, A. D. (1997). Service quality and operating efficiency synergies for management control in the provision of financial services: Evidence from Greek bank branches. *European Journal of Operational Research*, 98:300–313.

Athanassopoulos, A. D. (1998). Nonparametric frontier models for assessing the market and cost efficiency of large-scale bank branch networks. *Journal of Money, Credit and Banking*, 30(2):172–192.

Athanassopoulos, A. D. (2000). An optimisation framework of the triad; service capabilities, customer satisfaction and performance. In Harker, P. T. and Zenios, S. A., editors, *Performance of Financial Institutions; Efficiency, Innovation and Regulation*, pages 312–335. Cambridge University Press, Cambridge, UK.

Athanassopoulos, A. D. and Giokas, D. (2000). The use of data envelopment analysis in banking instaitutions: evidence from the commercial bank of Greece. *Interfaces*, 30(March-April):81–95.

Athanassopoulos, A. D., Soteriou, A. C., and Zenios, S. A. (2000). Disentangling within- and between-country efficiency differences of bank branches. In Harker, P. T. and Zenios, S. A., editors, *Performance of Financial Institutions; Efficiency, Innovation and Regulation*, pages 336–363. Cambridge University Press, Cambridge, UK.

Athanassopoulos, A. D. and Thanassoulis, E. (1995). Separating market efficiency from profitability and its implications for planning. *Journal of the Operational Research Society*, 46(1):30–45.

Avkiran, N. K. (1994). Developing an instrument to measure customer service quality in branch banking. *International Journal of Bank Marketing*, 12(6):10–18.

Avkiran, N. K. (1997). Models of retail performance for bank branches: predicting the level of key business drivers. *International Journal of Bank Marketing*, 15(6):224–237.

Avkiran, N. K. (1999a). An application reference for data envelopment analysis in branch banking: helping the novice researcher. *International Journal of Bank Marketing*, 17(5):206–220.

Avkiran, N. K. (1999b). Quality customer service demands human contact. *International Journal of Bank Marketing*, 17(2):61–71.

Balk, B. M. (2001). Scale efficiency and productivity change. *Journal of Productivity Analysis*, 15(3):159–183.

Banker, R. D. (1984). Estimating most productive scale size using data envelopment analysis. *European Journal of Operational Research*, 17:35–44.

Banker, R. D., Chang, H., and Cooper, W. W. (1996a). 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., Chang, H., and Cooper, W. W. (1996b). Simulation studies of efficiency, returns to scale and misspecification with nonlinear functions in DEA. *Annals of Operations Research*, 66:233–253.

Banker, R. D., Charnes, A., and Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30:1078–1092.

Banker, R. D., Charnes, A., Cooper, W. W., and Maindiratta, A. (1988). A comparison of DEA and translog estimates of production frontiers using simulated observations from a known technology. In Dogramaci, A. and Färe, R., editors, *Applications of Modern production theory: efficiency and productivity*, pages 33–55. Kluwer Academic Publishers.

Banker, R. D., Conrad, R. F., and Strauss, R. P. (1986). A comparative application of data envelopment analysis and translog methods: An illustrative study of hospital production. *Management Science*, 32(1):30–43.

Banker, R. D., Gadh, V., and Gorr, W. (1993). A monte carlo comparison of two production frontier estimation methods: corrected ordinar least squares and data envelopment analysis. *European Journal of Operational Research*, 67:332–343.

Banker, R. D. and Maindiratta, A. (1986). Piecewise log-linear estimation of efficient production surfaces. *Management Science*, 32(1):126–135.

Banker, R. D. and Maindiratta, A. (1988). Nonparametric analysis of technical and allocative efficiencies in poduction. *Econometrica*, 56(6):1315–1332.

Banker, R. D. and Morey, R. C. (1986a). Efficiency analysis for exogenously fixed inputs and outputs,. *Operations Research*, 34(4):513–520.

Banker, R. D. and Morey, R. C. (1986b). The use of categorical variables in data envelopment analysis. *Management Science*, 32(12):1613–1627.

Banker, R. D. and Thrall, R. M. (1992). Estimation of returns to scale using data envelopment analysis. *European Journal of Operational Research*, 62:74–84.

Barber, C. B., Dobkin, D. P., and Huhdanpaa, H. (1996). The quickhull algorithm for convex hulls. *ACM Transcations on Mathematical Software*, 22(4):469–483. see also, http://www.geom.umn.edu/software/qhull (last visited in December 2001).

Bardhan, I., Bowlin, W. F., Cooper, W. W., and Sueyoshi, T. (1996). Models and measures for effcieicncy dominance in DEA. part i: Additive models and MED measures. *Journal of the Operations Research Society of Japan*, 39(3):322–332.

Barr, R. D., Seiford, L. M., and Siems, T. F. (1993). An envelopment-analysis approach to measuring the managerial efficiency of banks. *Annals of Operations Research*, 45:1–19.

Bauer, P. W., Berger, A. N., and Humphrey, D. B. (1993). Efficiency and productivity growth in U.S. banking. In Fried, H. O., Lovell, C. A. K., and Schmidt, S. S., editors, *The measurement of productive efficiency: Techniques and Applications*, pages 386–413. Oxford University Press, New York, Oxford.

REFERENCES

Berg, A. S., Forsund, F. R., and Jansen, E. S. (1991). Technical efficiency of Norwegian banks: The non-parametric approach to efficiency measurement. *Journal of Productivity Analysis*, 2(2):127–142.

Berg, S. A., Førsund, F. R., and Jansen, E. S. (1992). Malmquist indices of productivity growth during the deregulation of Norwegian banking, 1980-89. *Scandinavian Journal of Economics*, 94:S211–S228.

Berger, A. N. (1993). 'Distribution-free' estimates of efficiency in the U.S. banking industry and testes of the standard distributional approaches. *Journal of Productivity Analysis*, 4:261–292.

Berger, A. N., Hancock, D., and Humphrey, D. B. (1993). Bank efficiency derived from the profit function. *Journal of Banking and Finance*, 17:314–347.

Berger, A. N., Hanweck, G. A., and Humphrey, D. B. (1987). Competitive viability in banking: Scale, acope and product mix economies. *Journal of Monetary Economics*, 20:501–520.

Berger, A. N. and Humphrey, D. B. (1991). The dominance of inefficiencies over scale and product mix economies and banking. *Journal of Monetary Economics*, 28:117–148.

Berger, A. N. and 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., Leusner, J. H., and Mingo, J. J. (1994). The efficiency of bank branches. The Wharton Financial Institutions Center, Working Paper 94-27.

Berger, A. N., Leusner, J. H., and Mingo, J. J. (1997). The efficiency of bank branches. *Journal of Monetary Economics*, 40(1):141 – 162.

Berger, A. N. and Mester, L. J. (2000). Inside the black box; what explains differences in the efficiencies of financial institutions? In Harker, P. T. and Zenios, S. A., editors, *Performance of Financial Institutions; Efficiency, Innovation and Regulation*, pages 93–150. Cambridge University Press, Cambridge, UK.

Bjurek, H. (1996). The Malmquist total factor productivity index. *Scandinavian Journal of Economics*, 98(2):303–314.

Bjurek, H., Førsund, F. R., and Hjalmarsson, L. (1998). Malmquist productivity indexes: An empirical comparison. In Färe, R., Grosskopf, S., and Russell, R. R., editors, *Index Numbers: Essays in Honour of Sten Malmquist*, pages 217–239. Kluwer Academic Publishers.

Bogetoft, P. (1996). DEA on relaxed convexity assumptions. *Management Science*, 42(3):457–465.

Bogetoft, P. and Hougaard, J. L. (1998). Efficiency evaluations based on potential (non-proportional) improvements. *Journal of Productivity Analysis*, 12:233–247.

Bogetoft, P. and Hougaard, J. L. (2001). Rational inefficiencies. Paper presented at the Exclusive DEA Workshop, Odense, Copenhagen, 21-22 September 2001.

Bogetoft, P., Tama, J. M., and Tind, J. (2000). Convex input and output projections of nonconvex production possibility sets. *Management Science*, 46:858–869.

Borges, P. C. (2000). CHESS:Changing Horizon Efficient Set Search. A simple principle for multiobjective optimisation. *Journal of Heuristics*, 6:405–418.

Boufounou, P. V. (1995). Evaluating bank branch location and performance: A case study. *European Journal of Operational Research*, 87:389–402.

Briec, W. (1998). Hölder distance function and measurement of technicall efficiency. *Journal of Productivity Analysis*, 11(2):111–131.

Brockett, P. L., Charnes, A., Cooper, W. W., Huang, Z. M., and Sun, D. B. (1997a). Data transformations in DEA cone ratio envelopment approaches for monitoring bank performances. *European Journal of Operational Research*, 98:250–268.

Brockett, P. L., Rousseau, J. J., Wang, Y., and Zhow, L. (1997b). Implementation of DEA models using GAMS. Research Report 765, University of Texas, Austin.

Camanho, A. S. and Dyson, R. G. (1999). Efficiency, size, benchmarks and targets for bank branches: An application of data envelopment analysis. *Journal of the Operational Research Society*, 50(9):903–915.

Canals, J. (1995). *Universal Banking*. Clarendon Press, Oxford.

Caves, D., Chistensen, L. R., and Diewert, W. (1982). The economic theory aof index numbers and the measurement of input, output, and productivity. *Econometrica*, 50(6):1393–1414.

Chambers, R. G., Chung, Y., and Färe, R. (1996a). Benefit and distance functions. *Journal of Economic Theory*, 70:407–419.

Chambers, R. G., Chung, Y., and Färe, R. (1998). Profit, directional distance functions, and Nerlovian efficiency. *Journal of Optimization Theory and Applications*, 98(2):351–364.

Chambers, R. G., Färe, R., and Grosskopf, S. (1996b). Productivity growth in APEC countries. *Pacific Economic Review*, 1(3):181–190.

Chambers, R. G. and Mitchell, T. (2001). Homotheticity and non-radial changes. *Journal of Productivity Analysis*, 15:31–19.

Chang, K.-P. and Guh, Y.-Y. (1991). Linear production functions and the data envelopment analysis. *European Journal of Operational Research*, 52:215–223.

Charnes, A., Clark, C., Cooper, W., and Golany, B. (1985a). A developmental study of data envelopment analysis in measuring the efficiency of maintainance units in the US air forces. *Annals of Operations Research*, 2:95–112.

Charnes, A., Cooper, W. W., Golany, B., Seiford, L., and Stutz, J. (1985b). Foundations of data envelopment analysis for Pareto-Koopmans efficient empirical production functions. *Journal of Econometrics*, 30:91–107.

Charnes, A., Cooper, W. W., Huang, Z. M., and Sun, D. B. (1990). Polyhedral cone-ratio DEA models with an illustrative application to large industrial banks. *Journal of Econometrics*, 46:73–91.

Charnes, A., Cooper, W. W., Lewin, A. Y., and Seiford, L. W. (1994). *Data Envelopment Analysis: Theory, Methodology and Applications*. Kluwer Academic Publishers, Boston.

Charnes, A., Cooper, W. W., and Rhodes, E. (1978). Measuring efficiency of decision making units. *European Journal of Operational Research*, 2:429–444.

Charnes, A., Cooper, W. W., and Rhodes, E. (1981). Evaluating program and managerial efficiency: An application of data envelopment analysis to program follow through. *Management Science*, 27(6):668–697.

Charnes, A., Haag, S., Jaska, P., and Semple, J. (1992). Sensitivity of efficiency classifications in the additive model of data envelopment analysis. *International Journal of Systems Science*, 23:789–798.

Chase, R. B. and Heskett, J. L. (1995). Introduction to the focused issue on service management. *Management Science*, 41(11):1717–1719.

Chavas, J.-P. and Cox, T. L. (1999). A generalized distance function and the analysis of production efficiency. *Southern Economic Journal*, 66(2):294–348.

Chen, Y. and Ali, A. I. (2002). Output-input ratio analysis and DEA frontier. *European Journal of Operational Research*, 142:476–479.

Cherchye, L., Kuosmanen, T., and Post, T. (2001). FDH directional distance functions with an aplication to European commercial banks. *Journal of Productivity Analysis*, 15(3):201–215.

Cherchye, L. and Puyenbroeck, T. V. (2001a). A comment on multi-stage DEA methodology. *Operations Research Letters*, 28:93–98.

Cherchye, L. and Puyenbroeck, T. V. (2001b). Product mixes as objects of choice in non-parametric efficiency measurement. *European Journal of Operational Research*, 132(2):287–295.

Cherchye, L. and Van Puyenbroeck, T. (1999a). Learning from input-output mixes in DEA: a proportional measure for slack-based efficient projections. *Managerial and Decision Economics*, 20:151–161.

Cherchye, L. and Van Puyenbroeck, T. (1999b). Non-radial efficiency and semi-radial efficiency. In Westermann, G., editor, *Data Envelopment Analysis in the Service Sector*, pages 51–64. Gabler Edition Wissenschaft.

Cherchye, L. and Van Puyenbroeck, T. (2003). profit efficiency analysis under limited information with an application to German farm types. NEP - New Economics Papers, Issue: nep-eff-2003-03-25.

Chilingerian, J. A. (1995). Evaluating physician efficiency in hospitals: a multivariate analysis of best practices. *European Journal of Operational Research*, 80:548–574.

Chung, Y., Färe, R., and Grosskopf, S. (1997). Productivity and undesirable outputs: a directional distance function approach. *Journal of Environmental Management*, 51(3):229–240.

Coelli, T. (1998). A multi-stage methodology for the solution of orientated DEA models. *Operations Research Letters*, 23:143–149.

Coelli, T., Grifell-Tatjé, and Perelman, S. (2002). Capacity utilisation and profitability: A decomposition of short run profit efficiency. *International Journal of Production Economics*, 79:261–278.

Coelli, T., Perelman, S., and Romano, E. (1999). Accounting for environmental influences in stochastic frontier models: with application to international airlines. *Journal of Productivity Analysis*, 11(2):251–273.

Coelli, T., Rao, D. S. P., and Battese, G. E. (1998). *An Introduction to Efficiency and Productivity Analyisis*. Kluwer Academic Publishers, Boston/Dordrecht/London.

Colwell, R. J. and Davis, E. P. (1992). Output and productivity in banking. *Scandinavian Journal of Economics*, 94(Supplement):S111–S129.

Cook, W. D. and Hababou, M. (2001). Sales performance measurement in bank branches. *Omega, The International Journal of Management Science*, 29:299–307.

Cook, W. D., Hababou, M., and Tuenter, H. J. H. (2000). Multicomponent efficiency measurement and shared inputs in data envelopment analysis: an application to sales and service performance in bank branches. *Journal of Productivity Analysis*, 14:209–224.

Cook, W. D., Kazakov, A., and Roll, Y. (1994). On the measurement and monitoring of relative efficiency of highway maintenance patrols. In Charnes, A., Cooper, W. W., Lewin, A. Y., and Seiford, L. W., editors, *Data Envelopment Analysis, Theory, Methodology and Applications*, pages 195–210. Kluwer Academic Publishers.

Cooper, W., Park, K. S., and Pastor, J. T. (2001). The range adjusted measure (RAM) in DEA: A response to the comment by Steinmann and Zweifel. *Journal of Productivity Analysis*, 15(2):145–152.

Cooper, W. W., Park, K. S., and Pastor, J. T. (1999). RAM: A range measure of inefficiency for use with additive models, and relations to other models and measures in DEA. *Journal of Productivity Analysis*, 11:5–42.

Cooper, W. W., Seiford, L. M., and Tone, K. (2000). *Data Envelopment Analysis: A comprehensive text with models, applications, references and DEA-Solver software*. Kluwer Academic Publishers, Boston.

Cooper, W. W., Thompson, R. G., and Thrall, R. M. (1996). Introduction: Extensions and new developments in DEA. *Annals of Operations Research*, 66:3–45.

De Borger, B. and Kerstens, K. (1996). Radial and nonradial measures of technical efficiency: An empirical illustration for Belgium local governments using an FDH reference technology. *Journal of Productivity Analysis*, 7:41–62.

De Young, R. and Nolle, D. (1996). Foreign-owned banks in the US: Earning market share or buying it? *Journal of Money, Credit, and Banking*, 28:622–636.

Dekker, D. and Post, T. (2001). A quasi-concave DEA model with an application for bank branch performance evaluation. *European Journal of Operational Research*, 132(2):296–311.

Deprins, D., Simar, L., and Tulkens, H. (1984). Measuring labour efficiency in post-offices. In Marchand, M., Pestieau, P., and Tulkens, H., editors, *The performance of Public Enterprises: Concepts and Measurement*, pages 243–267. Elsevier Science Publishers B.V., Amsterdan, North Holland.

Dervaux, B., Kerstens, K., and Vanden-Eeckaut, P. (1998). Radial and nonradial static efficiency decompositions: A focus on congestion. *Transportation Research*, 32B(5):299–312.

Diewert, E. R. and Nakamura, A. O. (2003). Index number concepts, measures and decompositions of productivity growth. *Journal of Productivity Analysis*, 19(2/3):127–159.

Doukas, J. and Switzer, L. N. (1991). Economies of scale and scope in Canadian branch banking. *Journal of International Financial Markets Institutions and Money*, 1:61–84.

Drake, L. and Howcroft, B. (1994). Relative efficiency in the branch network of a UK bank: An empirical study. *Omega: The International Journal of Management Science*, 22(1):83–90.

Drake, L. and Howcroft, B. (1995). Measuring the relative efficiency of the selling function: An application of data envelopment analysis to UK bank branches. Loughborough University Banking Centre, Working Paper N. 89/95.

Dusansky, R. and Wilson, P. W. (1994). Technical efficiency in the decentralized care of the developmentally disabled,. *Review of Economics and Statistics*, 76:340–345.

Dusansky, R. and Wilson, P. W. (1995). On the relative efficiency of alternative models of producing a public sector output: The case of the developmentally disabled. *European Journal of Operational Research*, 80:608–618.

Dyckhoff, H. and Allen, K. (2001). Measuring ecological efficiency with data envelopment analysis. *European Journal of Operational Research*, 132(2):312–325.

Dyson, R. G. and Thanassoulis, E. (1988). Reducing weight flexibility in data envelopment analysis. *Journal of the Operational Research Society*, 39(6):563–576.

ECB (1999a). The effects of technology on the EU banking systems. European Central Bank, July.

ECB (1999b). Possible effects of EMU on the EU banking systems in the medium and long term. European Central Bank, February.

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society*, Series A, general 120(Part 3):253–281.

Faulhaber, G. R. (1995). Banking markets; productivity, risk, and customer satisfaction. Working paper no. 95-14, The Warthon Financial Institutions Center.

Ferrier, G. D. and Lovell, C. A. K. (1990). Measuring cost efficiency in banking: Econometric and linear programming evidence. *Journal of Econometrics*, 46:229–245.

Färe, R., Grifell-Tatjé, E., Grosskopf, S., and Lovell, C. A. K. (1997a). Biased technical change and malmquist productivity index. *Scandinavian Journal of Economics*, 99(1):119–128.

Färe, R. and Grosskopf, S. (1994a). *Cost and Revenue Constrained Production*. Springer-Verlag, New York.

Färe, R. and Grosskopf, S. (1994b). Estimation of returns to scale using data envelopment analysis :a comment. *European Journal of Operational Research*, 79:379–382.

Färe, R., Grosskopf, S., and Lee, H. (1990). A nonparametric approach to expenditure constrained profit maximization. *American Journal of Agricultural Economics*, 12(3):574–581.

Färe, R., Grosskopf, S., and Lee, W.-F. (2001). Productivity and technical change: the case of Taiwan. *Applied Economics*, 33(15):1911–1925.

Färe, R., Grosskopf, S., and Lovell, A. K. (1994a). *Production Frontiers*. Cambridge University Press, UK.

Färe, R., Grosskopf, S., and Lovell, C. A. K. (1985). *The Measurement of Efficiency of Production*. Kluwer-Nijhoff Publishing, Boston.

Färe, R., Grosskopf, S., and Norris, M. (1997b). Productivity growth, technical progress and efficiency changes in industrialized countries: Reply. *The American Economic Review*, 87(5):1040–1043.

Färe, R., Grosskopf, S., Norris, M., and Zhang, Z. (1994b). Productivity growth, technical progress and efficiency changes in industrialised countries. *American Economic Review*, 84:66–83.

Färe, R., Grosskopf, S., and Roos, P. (1998). Malmquist productivity indexes: A survey of theory and practice. In Färe, R., Grosskopf, S., and Russell, R. R., editors, *Index Numbers: Essays in Honour of Sten Malmquist*, pages 127–190. Kluwer Academic Publishers.

Färe, R., Grosskopf, S., and Weber, W. L. (1989). Measuring school district performance. *Public Finance Quarterly*, 17(4):409–428.

Färe, R. and Li, S. K. (1998). Inner and outer approximations of technology: a data envelopment analysis approach. *European Journal of Operational Research*, 105:622–625.

Färe, R. and Lovell, C. A. K. (1978). Measuring the technical efficiency of production. *Journal of Economic Theory*, 19(1):150–162.

Färe, R. and Primont, D. (1995). *Multi-output production and duality: Theory and applications*. Kluwer academic publishers, Boston, London, Dordrecht.

Frei, F. X. and Harker, P. T. (1999). Projections onto efficient frontiers: theoretical and computational extensions to DEA. *Journal of Productivity Analysis*, 11(3):275–300.

Frei, F. X., Harker, P. T., and Hunter, L. W. (2000). Inside the black box: What makes a bank efficient? In Harker, P. T. and Zenios, S. A., editors, *Performance of Financial Institutions; Efficiency, Innovation and Regulation*, pages 259–311. Cambridge University Press, Cambridge, UK.

Fried, H. O., Schmidt, S. S., and Yaisawarng, S. (1999). Incorporating the operational environment into a nonparametric measure of technical efficiency. *Journal of Productivity Analysis*, 12:249–267.

Førsund, F. R. (1993). Productivity growth in Norwegian ferries. In Fried, H. O., Lovell, C. A. K., and Schmidt, S. S., editors, *The measurement of productive efficiency: Techniques and Applications*, pages 10000–159. Oxford University Press, New York, Oxford.

Førsund, F. R. (1996). On the calculation of the scale elasticity in DEA models. *Journal of Productivity Analysis*, 7:283–302.

Førsund, F. R. (1998). The rise and fall of slacks: Comments on quasi-malmquist productivity indices. *Journal of Productivity Analysis*, 10:21–34.

Førsund, F. R. (2001). Categorical variables in DEA. ICER Working Paper.

Førsund, F. R. and Hernaes, E. (1994). A comparative analysis of ferry transport in Norway. In Charnes, A., Cooper, W. W., Lewin, A. Y., and Seiford, L. W., editors, *Data Envelopment Analysis, Theory, Methodology and Applications*, pages 285–311. Kluwer Academic Publishers.

Førsund, F. R. and Hjalmarsson, L. (1979). Generalised Farrell measures of efficiency: an application to milk processing in Swedish dairy plants. *The Economic Journal*, 89(June):294–315.

Førsund, F. R., Lovell, C. A. K., and Schmidt, P. (1980). A survey of frontier production functions and of their relationship to efficiency measurement. *Journal of Econometrics*, 13:5–25.

Fukuyama, H. (2000). Returns to scale and scale elsticity in data envelopment analysis. *European Journal of Operational Research*, 125(1):93–112.

Fukuyama, H. (2003). Scale characterizations in a DEA directional technology distance function framework. *European Journal of Operational Research*, 144(1):108–127.

Giokas, D. (1991). Bank branch operating efficiency: A comparative application of DEA and the loglinear model. *Omega: The International Journal of Management Science*, 19(6):549–557.

Golany, B., Phillips, F., and Rousseau, J. (1993). Models for improved effectiveness based on DEA efficiency results. *IIE Transactions*, 25(6):2–10.

Golany, B. and Roll, Y. (1993). Some extensions of techniques to handle non-discretionary factors in data envelopment analysis. *Journal of Productivity Analysis*, 4:419–432.

Golany, B. and Roll, Y. (1994). Incorporating standards via DEA. In Charnes, A., Cooper, W. W., Lewin, A. Y., and Seiford, L. W., editors, *Data Envelopment Analysis, Theory, Methodology and Applications*, pages 313–328. Kluwer Academic Publishers.

Golany, B. and Storbeck, J. E. (1999). A data envelopment analysis of the operational efficiency of bank branches. *Interfaces*, 29(3):14–26.

Golany, B. and Yu, G. (1997). Estimating returns to scale in DEA. *European Journal of Operational Research*, 103:28–37.

Gold, B. (1981). Changing perspectives on size, scale, and returns: An interpretative survey. *Journal of Economic Literature*, XIX:5–33.

Gong, B.-H. and Sickles, R. (1992). Finite sample evidence on the performance of stochastic frontiers and data envelopment analysis using panel data. *Journal of Econometrics*, 51:259–284.

Green, R. H., Cook, W., and Doyle, J. (1997). A note on the additive data envelopment analysis model,. *Journal of the Operational Research Society*, 48:446–448.

Green, R. H., Doyle, J. R., and Cook, W. D. (1996). Efficiency bounds in data envelopment analysis. *European Journal of Operational Research*, 89:482–490.

Greene, W. H. (1993). The econometric approach to efficiency analysis. In Fried, H. O., Lovell, C. A. K., and Schmidt, S. S., editors, *The measurement of productive efficiency: Techniques and Applications*, pages 68–119. Oxford University Press, New York, Oxford.

Grifell-Tatjé, E. and Lovell, C. A. K. (1995). A note on the Malmquist productivity index. *Economics Letters*, 47(2):169–175.

Grifell-Tatjé, E., Lovell, C. A. K., and Pastor, J. T. (1998). A quasi-malmquist productivity index. *Journal of Productivity Analysis*, 10:7–20.

Grönroos, C. (2000). *Service Management and Marketing, a customer relationship management approach.* John Wiley and Sons, LTD, second edition.

Grosskopf, S. (1996). Statistical inference and nonparametric efficiency: A selective survey. *Journal of Productivity Analysis*, 7:161–176.

Haag, S. E. and Jaska, P. V. (1995). Interpreting inefficiency ratings: an application of bank branch operating efficiencies. *Managerial and Decision Economics*, 16:7–14.

Habenicht, W. (1982). Quad Trees, a datastructure for discrete vector optimization problems. *Lecture Notes in Economics and Systems*, 209:136–145.

Hair JR, J. F., Anderson, R. E., Tatham, R. L., and Black, W. C. (1995). *Multivariate Data Analysis with Readings.* Prentice Hall International Editions, fourth edition.

Halme, M., Joro, T., and Koivu, M. (2002). Dealing with interval data in data envelopment analysis. *European Journal of Operational Research*, 137:22–27.

Hartman, T. E., Storbeck, J. E., and Byrnes, P. (2001). Allocative efficiency in branch banking. *European Journal of Operational Research*, 134:232–242.

Heskett, J. L., Jones, T. O., Loveman, G. W., Sasser, W. E., and Schlesinger, L. A. (1994). Putting the service-profit chain to work. *Harvard Business Review*, March-April:164–174.

Hjalmarsson, L., Kumbhakar, S. C., and Heshmati, A. (1996). DEA, DFA and SFA: A comparison. *Journal of Productivity Analysis*, 7:303–327.

Howcroft, J. B. (1989). Retail branch banking issues in the United Kingdom. *Journal of Retail Banking*, XI(1):11–17.

Howcroft, J. B. (1991). Customer satisfaction in retail banking. *Service Industries Journal*, 11(1):11–17.

Howcroft, J. B. (1992). Contemporary issues in UK bank delivery systems. *International Journal of Service Industry Management*, 3(1):39–56.

Howcroft, J. B. and Beckett, A. (1993). Change in the UK bank branch networks: A customer perspective. *The Service Industries Journal*, 13(4):267–288.

Howland, A. (2000). The evolution of the bank branch. *Canadian Banker*, 1st quarter:25–29.

Huang, Z., Li, S., and Rousseau, J. (1997). Determining rates of change in data envelopment analysis. *Journal of the Operational Research Society*, 48:591–599.

Humphrey, D. B. and Pulley, L. B. (1997). Bank's responses to deregulation; profits, technology, and efficiency. *Journal of Money, Credit, and Banking*, 29(1):73–93.

Kamakura, W. A. (1988). A note on the use of categorical variables in data envelopment analysis. *Management Science*, 34(10):1273–1276.

Kamakura, W. A., Lenartowicz, T., and Ratchfrord, B. T. (1996). Productivity assessment of multiple retail outlets. *Journal of Retailing*, 72(4):333–356.

Kantor, J. and Maital, S. (1999). Measuring efficiency by product group:integrating DEA with activity-based accounting in a large mideast bank. *Interfaces*, 29(3):27–36.

Kerstens, K. (1996). Technical efficiency measurement and explanation of French urban transit companies. *Transportation Research A*, 30(6):431–452.

Kerstens, K. and Vanden-Eeckaut, P. (1995). Technical efficiency measures on DEA and FDH: A reconsideration of the axiomatic literature. CORE discussion paper 9513, Center for Operations Research and Econometrics, Universite Catholique de Louvain, Belgium.

Kirjavainen, T. and Loikkanen, H. A. (1998). Efficiency differences of Finnish senior secondary schools: An application of DEA and Tobit analysis. *Economics of Education Review*, 17(4):377–394.

Klassen, K. J., Russel, R. M., and Chrisman, J. J. (1998). Efficiency and productivity measures for high contact serviçes. *The Service Industries Journal*, 18(4):1–18.

Knox, K. J., Blankmeyer, E., and Stutzman, J. R. (1999). Relative economic efficiency in Texas nursing facilities: A profit function analysis. *Journal of Economics and Finance*, 23(3):199–213.

Kordupleski, R. E., Rust, R. T., and Zahorik, A. J. (1993). Why improving quality doesn't improve quality (or whatever happened to marketing?). *California Management Review*, 35(Spring):82–95.

Krivonozhko, V. E., Utkin, O. B., Volodin, A. V., and Sablin, I. A. (2001). Application of DEA approach to production units with some negative outputs. Paper presented at OR43, University of Bath, UK, 4-6 September.

Kumbhakar, S. C. (1996). A parametric approach to efficiency measurement using a flexible profit funtion. *Southern Economic Journal*, 63(2):473–487.

Kumbhakar, S. C. (2001). Estimation of profit functions when profit is not maximum. *American Journal of Agricultural Economics*, 83(1):1–19.

Kuosmanen, T. (1999a). Data envelopment analysis of non-convex technology:with empirical evidence from the batting technology of Finnish super-league pesis players. Working paper 224, Helsinki School of economics and business administration.

Kuosmanen, T. (1999b). Some remarks on scale efficiency and returns to scale in DEA. Helsinki School of economics and business administration.

Kuosmanen, T. (2001). DEA with efficiency classification preserving conditional convexity. *European Jounal of Operational Research*, 132(2):326–342.

Kuosmanen, T. and Post, T. (2001). Measuring economic efficiency with incomplete price information with an application to European commercial banks. *European Journal of Operational Research*, forthcoming.

Kuosmanen, T. and Post, T. (2002). Quadratic data envelopment analysis. *Journal of the Operational Research Society*, 53(11):1204–1214.

Lau, L. J. and Yotopoulos, P. A. (1971). A test for relative efficiency and application to indian agriculture. *The American Economic Review*, 61:94–109.

Lewis, B. R. (1989). Quality in the service sector: A review. *International Journal of Bank Marketing*, 7(5):4–12.

Lovell, C. A. K. (1993). Production frontiers and productive efficiency. In Fried, H. O., Lovell, C. A. K., and Schmidt, S. S., editors, *The measurement of productive efficiency: Techniques and Applications*, pages 3–67. Oxford University Press, New York, Oxford.

Lovell, C. A. K. (1995). Measuring the macroeconomic performance of the Taiwanese economy. *International Journal of Production Economics*, 39:165–178.

Lovell, C. A. K. (2001). The decomposition of Malmquist productivity indexes. Paper presented at the 7th European Workshop on Efficiency and Productivity Analysis, Oviedo, Spain, 25-29 September.

Lovell, C. A. K. and Pastor, J. T. (1995). Units invariant and translation invariant DEA models. *Operations Research Letters*, 18:147–151.

Lovell, C. A. K. and Pastor, J. T. (1997). Target setting: An application to a bank branch network. *European Journal of Operational Research*, 98:290–299.

Lovell, C. A. K. and Schmidt, P. (1988). A comparison of alternative approaches to the measurement of productive efficiency. In Dogramaci, A. and Färe, R., editors, *Applications of Modern production theory: efficiency and productivity*, pages 3–32. Kluwer Academic Publishers, Boston, Dordrecht, Lancaster.

Lovell, C. A. K. and Sickles, R. C. (1983). Testing efficiency hypothesis in joint production: a parametric approach. *Review of Economics and Statistics*, 65(1):51–58.

Lovelock, C. (1996). *Services Marketing*. Prentice Hall International Editions, 3rd edition.

Loveman, G. W. (1998). Employee satisfaction, customer loyalty, and financial performance: An empirical examination of the service profit chain in retail banking. *Journal of Service Research*, 1(1):18–31.

Lozano Vivas, A. (1997). Profit efficiency for Spanish savings banks. *European Journal of Operational Research*, 98:381–394.

Löthgren, M. and Tambour, M. (1996). Alternative approaches to estimate returns to scale in DEA-models. Working paper series in economics and finance 90, Stockholm School of Economics.

Manandhar, R. and Tang, J. C. S. (2001). The evaluation of bank branch performance using data envelopment analysis. *Journal of High Technology Management Research*, 0:1–17.

McCarty, T. A. and Yaisawarng, S. (1993). Technical efficiency in New Jersey school districts. In Fried, H. O., Lovell, C. A. K., and Schmidt, S. S., editors, *The measurement of productive efficiency: Techniques and Applications*, pages 271-287. Oxford University Press, New York, Oxford.

Molineux, P., Altunbas, Y., and Gardener, E. (1996). *Efficiency in European Banking*. John Wiley and Sons, LTD.

Muñiz, M. A. (2002). Separating managerial inefficiency and external conditions in data envelopment analysis. *European Journal of Operational Research*, 143:625–643.

Mukherjee, A., Nath, P., and Pal, M. (2003). Resource, service quality and performance triad: a framework for measuring the efficiency of banking services. *Journal of the Operational Research Society*, 54(7):723–735.

Murphy, N. B. and Orgler, Y. E. (1982). Cost analysis for branching systems: methodology, test results and implications for management. *Journal of Financial Research*, 5:181–188.

Nash, D. and Karwat-Sterna, A. (1996). An application of DEA to measure branch cross selling efficiency. *Computers and Operations Research*, 23(4):385–392.

Olesen, O. B. and Petersen, N. C. (1996). Indicators of ill-conditioned data sets and model misspecification in Data Envelopment Analysis: An extended facet approach. *Management Science*, 42(2):205–219.

Olesen, O. B. and Petersen, N. C. (2002). Identification and use of efficient faces and facets in DEA. Forthcoming in Journal of Productivity Analysis.

Ondrich, J. and Ruggiero, J. (2002). Outlier detection in data envelopment analysis: an analysis of jackknifing. *Journal of the Operational Research Society*, 53(3):342–346.

Oral, M., Kettani, O., and Yolalan, R. (1992). An empirical study on analysing the produtivity of bank branches. *IIE Transactions*, 24(5):166–176.

Oral, M. and Yolalan, R. (1990). An empirical study on measuring operating efficiency and profitability of bank branches. *European Journal of Operational Research*, 46(3):282–294.

Orea, L. (2002). Parametric decomposition of a generalizes Malmquist productivity index. *Journal of Productivity Analysis*, 18(1):5–22.

Parasuraman, A., Zeithaml, V. A., and Berry, L. L. (1985). A conceptual model of service quality and its implications for future research. *Journal of Marketing*, 49(Fall):41–50.

Parasuraman, A., Zeithaml, V. A., and Berry, L. L. (1988). Servqual: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, 64(1):12–40.

Parkan, C. (1987). Measuring the efficiency of service operations: An application to bank branches. *Engeneering Costs and Production Economics*, 12:237–242.

Pastor, J. T. (1994). How to discount environmental effects in DEA: an application to bank branches. Working paper N. 011/94, Depto. De Estadistica e Investigacion Operativa, Universidad de Alicante, Spain.

Pastor, J. T. (1996). Translation invariance in data envelopment analysis: A generalisation. *Annals of Operations Research*, 66:93–102.

Pastor, J. T., Ruiz, J. L., and Sirvent, I. (1999). An enhanced DEA Russell graph efficiency measure. *European Journal of Operational Research*, 115:596–607.

Pavlopoulos, P. G. and Kouzelis, A. K. (1989). Cost behaviour in the banking industry: evidence from a Greek commercial bank. *Applied Economics*, 21:285–293.

Petersen, N. C. (1990). Data envelopment analysis on a relaxed set of assumptions. *Management Science*, 36(3):305–314.

Podinovski, V. V. (1999). Side effects of absolute weight bounds in DEA models. *European Journal of Operational Research*, 115:583–595.

Podinovski, V. V. (2001). DEA models for the explicit maximisation of relative efficiency. *European Journal of Operational Research*, 131:572–586.

Podinovski, V. V. and Athanassopoulos, A. (1998). Assessing the relative efficiency of decision making units using DEA models with weights restrictions. *Journal of the Operational Research Society*, 49(5):500–508.

Portela, M. C. A. S., Borges, P., and Thanassoulis, E. (2003). Finding closest targets in non-oriented DEA models: The case of convex and non-convex technologies. *Journal of Productivity Analysis*, 19(2/3):251–269.

Portela, M. C. A. S. and Thanassoulis, E. (2001). Decomposing school and school type efficiency. *European Journal of Operational Research*, 132(2):114–130.

Post, T. (1999). Estimating non-convex production sets. imposing convex input sets and output sets in Data Envelopment Analysis. Paper presented at the Sixth European Workshop on Efficiency and Productivity Analysis, Copenhaguen.

Post, T. (2001a). Estimating non-convex production sets - imposing convex input sets and output sets in data envelopment analysis. *European Journal of Operational Research*, 131(1):132–142.

Post, T. (2001b). Transconcave data envelopment analysis. *European Jounal of Operational Research*, 132(2):374–389.

Py, B. (1990). *Statistique Descriptive.* Economica, Paris, 3e Édition edition.

Radecki, L. J., Wenninger, J., and Orlow, D. K. (1996). Bank branches in supermarkets. *Current Issues in Economics and Finance*, 2(13):1–6.

Ray, S. C. (1988). Data envelopment analysis, nondiscretionary inputs and efficiency: an alternative interpretation. *Socio-Economic Planning Sciences*, 22(4):167–176.

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

Ray, S. C. (2001). On an extended decomposition of the Malmquist productivity index. Paper presented at the 7th European Workshop on Efficiency and Productivity Analysis, Oviedo, Spain, September 25-29.

Ray, S. C. and Desli, E. (1997). Productivity growth, technical progress and efficiency changes in industrialized countries: Comment. *The American Economic Review*, 87(5):1033–1039.

Reichheld, F. F. and Sasser, W. E. J. (1990). Zero defections: Quality comes to services. *Harvard Business Review*, Sept/Oct:103–109.

Roll, Y., Cook, W. D., and Golany, B. (1991). Controlling factor weights in data envelopment analysis. *IIE Transactions*, 23(1):2–9.

Roth, A. V. and Jackson III, W. E. (1995). Strategic determinants of service quality and performance: Evidence from the banking industry. *Management Science*, 41(11):1720–1733.

Rousseau, J. J. and Semple, J. H. (1993). Notes: Categorical outputs in data envelopment analysis. *Management Science*, 39(3):384–386.

Räty, T. (2002). Efficient facet based efficiency index: A variable returns to scale specification. *Journal of Productivity Analysis*, 17(1/2):65–82.

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

Ruggiero, J. (1998). Non-discretionary inputs in data envelopment analysis. *European Journal of Operational Research*, 111(3):461–469.

Ruggiero, J. and Bretschneider, S. (1998). The weighted Russel measure of technical efficiency. *European Journal of Operational Research*, 108:438–451.

Russell, R. R. (1985). Measures of technical efficiency. *Journal of Economic Theory*, 35(1):109–126.

Rust, R. T., Zahorik, A. J., and Keiningham, T. L. (1995). Return on quality (ROQ): Making service quality finacially accountable. *Journal of Marketing*, 59(April):58–70.

Schaffnit, C., Rosen, D., and Paradi, J. C. (1997). Best practice analysis of bank branches: An application of DEA in a large canadian bank. *European Journal of Operational Research*, 98:269–289.

Schneider, B. (1991). Service quality and profits: Can you have your cake and eat it, too? *Human Resource Planning*, 14(2):151–157.

Seiford, L. M. and Zhu, J. (1999). An investigation of returns to scale in data envelopment analysis. *Omega, The International Journal of Management Science*, 27:1–11.

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

Sevcovic, D., Halicka, M., and Brunovsky, P. (2002). DEA analysis for a large structured bank branch network. Forthcoming in Central European Journal of Operations Research.

Sherman, H. D. (1984). Improving the productivity of service business. *Sloan Management Review*, Spring:11–23.

Sherman, H. D. and Gold, F. (1985). Bank branch operating efficiency: Evaluation with DEA. *Journal of Banking and Finance*, 9(2):297–315.

Sherman, H. D. and Ladino, G. (1995). Managing bank productivity using data envelopment analysis (dea). *Interfaces*, 25(March-April):60–73.

Simar, L. and Wilson, P. W. (2003). Estimation and inference in two-stage, semi-parametric models of production processes. Tecnhical Report 0310, IAP Statistics Network.

Soteriou, A. and Stavrinides, Y. (2000). An internal customer service quality data envelopment analysis model for bank branches. *International Journal of Bank Marketing*, 18(5):246–252.

Soteriou, A. and Zenios, S. A. (1999). Operations, quality and profitability in the provision of banking services. *Management Science*, 45(9):1221–1238.

Soteriou, A. C. and Stavrinides, Y. (1997). An internal customer service quality data envelopment analysis model for bank branches. *International Journal of Operations and Production Management*, 17(8):780–789.

Staat, M. (1999). Treating non-discretionary variables one way or the other: implications for efficiency scores and their interpretation. In Westermann, G., editor, *Data Envelopment Analysis in the Service Sector*, pages 23–49. Gabler Edition Wissenschaft.

Steinmann, L. and Zweifel, P. (2001). The range adjusted measure (RAM) in DEA: Comment. *Journal of Productivity Analysis*, 15(2):139–144.

Thanassoulis, E. (1993). A comparison of regression analysis and data envelopment analysis as alternative methods for performance assessments. *Journal of Operational Research Society*, 44(11):1129–1144.

Thanassoulis, E. (1999a). Data envelopment analysis and its use in banking. *Interfaces*, 29(3):1–13.

Thanassoulis, E. (1999b). Setting achievement targets for school children. *Education Economics*, 7(2):101–119.

Thanassoulis, E. (2001). *Introduction to the theory and application of Data Envelopment analysis: A foundation text with intefrated software.* Kluwer Academic Publishers.

Thanassoulis, E. and Allen, R. (1998). Simulating weights restrictions in data envelopment analysis by means of unibserved DMUs. *Management Science*, 44(4):586–594.

Thanassoulis, E., Boussofiane, A., and Dyson, R. G. (1995). Exploring output quality targets in the provision of perinatal care in England using data envelopment analysis. *European Journal of Operational Research*, 80:588–607.

Thanassoulis, E. and Dyson, R. (1992). Estimating preferred input-output levels using data envelopment analysis. *European Journal of the Operational Research Society*, 56:80–97.

Thanassoulis, E., Dyson, R. G., , and Foster, M. J. (1987). Relative efficiency assessments using data envelopment analysis: An application to data on rates departments. *Journal of Operational Research Society*, 38(5):397–411.

Thanassoulis, E., Portela, M. C. A. S., and Allen, R. (2003). Incorporating value judgments in DEA. In Cooper, W. W., Seiford, L. W., and Zhu, J., editors, *DEA Handbook*. Kluwer Academic Publishers.

Thompson, R., Langemeier, L., Lee, C., Lee, E., and Thrall, R. M. (1990). The role of multiplier boundas in efficiency analysis with application to kansas farming. *Journal of Econometrics*, 46:93–108.

Thompson, R. and Thrall, R. (1994). Polyhedral assurance regions with linked constraints. In Cooper, W. and Whinston, A., editors, *New directions in computational economics*, pages 121–133. Kluwer Academic Publishers.

Thompson, R. G., Dharmapala, P. S., Rothenberg, L. J., and Thrall, R. M. (1996). DEA/AR efficiency and profitability of 14 major oil companies in US exploration and production. *Computers and Operations Research*, 23(4):357–373.

Thompson, R. G., Dharmapala, P. S., and Thrall, R. M. (1995). Linked-cone DEA profit ratios and technical efficiency with application to illinois coal mines. *International Journal of Production Economics*, 39:99–115.

Thompson, R. G., Singleton, J. F. D., Thrall, R. M., and Smith, B. A. (1986). Comparative site evaluations for locating a high-energy physics lab in Texas. *Interfaces*, 16:35–49.

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

Thrall, R. M. (2000). Measures in DEA with an application to the Malmquist index. *Journal of Productivity Analysis*, 13(2):125–137.

Tofallis, C. (1996). Improving discernment in DEA using profiling. *Omega, The International Journal of Management Science*, 24(3):361–364.

Tone, K. (1993). An e-free dea and a new measure of efficiency. *Journal of the Operations Research Society of Japan*, 36(3):167–174.

Tone, K. (2001). A slacks-based measure of efficiency in data envelopment analsysis. *European Journal of Operational Research*, 130:498–509.

Tone, K. and Sahoo, B. K. (2003). Scale, indivisibilities and production function in data envelopment analysis. *International Journal Production Economics*, 84:165–192.

Tortosa-Ausina, E. (2002). Bank cost efficiency and output specification. *Journal of Productivity Analysis*, 18(3):199–222.

Tulkens, H. (1993). On FDH efficiency analysis: Some methodological issues and applications to retail banking, courts and urban transit. *Journal of Productivity Analysis*, 4:183–210.

Tulkens, H. and Malnero, A. (1996). Nonparametric approaches to the assessment of the relative efficiency of bank branches. In Mayes, D. G., editor, *Sources of productivity growth*, pages 223–244. Cambridge University Press, Cambridge.

Tulkens, H. and Vanden-Eeckaut, P. (1995). Non-parametric efficiency, progress and regress measures for panel data: methodological aspects. *European Journal of Operational Research*, 80(3):474–499.

Varian, H. R. (1992). *Microeconomic analysis*. W.W. Norton and Company, 3rd edition.

Vassiloglou, M. and Giokas, D. (1990). A study of the relative efficiency of bank branches: An application of data envelopment analysis. *Journal of Operational Research Society*, 41(7):591–597.

Wheelock, D. C. and Wilson, P. W. (1999). Technical progress, inefficiency, and productivity change in U.S. banking, 1984-1993. *Journal of Money, Credit, and Banking*, 31(2):212–234.

Wilson, P. W. (1995). Detecting influential observations in data envelopment analysis. *Journal of Productivity Analysis*, 6:27–45.

Wong, Y. H. B. and Beasley, J. E. (1990). Restricting weight flexibility in data envelopment analysis. *Journal of Operational Research Society*, 41(9):829–835.

Xue, M. and Harker, P. T. (1999). Overcoming the inherent dependency of DEA efficiency scores: A bootstrap approach. Working paper 99-17 Wharton Financial Institutions Center.

Yu, G., Wei, Q., Brockett, P., and Zhou, L. (1996). Construction of all DEA efficient surfaces of the production possibility set under the generalized data envelopment analysis model. *European Journal of Operational Research*, 95:491–510.

Zardkoohi, A. and Kolari, J. (1994). Branch office economies of scale and scope: Evidence from savings banks in Finland. *Journal of Banking and Finance*, 18:421–432.

Zeithaml, V. A. (2000). Service quality, profitability, and the economic worth of customers: what we know and what we need to learn. *Journal of the Academy of Marketing Science*, 28(1):67–85.

Zeithaml, V. A., Berry, L. L., and Parasuraman, A. (1996). The behavioral consequences of service quality. *Journal of Marketing*, 60:31–46.

Zenios, C. V., Zenios, S. A., Agathocleous, K., and Soteriou, A. C. (1999). Benchmarks of the efficiency of bank branches. *Interfaces*, 29(3):37–51.

Zhang, Y. and Bartels, R. (1998). The effect of sample size on the mean efficiency in DEA with an application to electric distribution in Australia, Sweden and New Zeland. *Journal of Productivity Analysis*, 9:187–204.

Zhu, J. (2000). Setting scale efficient targets in DEA via returns to scale estimation method. *Journal of the Operational Research Society*, 51(3):376–378.

Zhu, J. and Shen, Z. (1995). A discussion of testing DMUs returns to scale. *European Journal of Operational Research*, 81:590–596.

Zieschang, K. (1984). An extended Farrell technical efficiency measure. *Journal of Economic Theory*, 33:387–396.

Zofio, J. L. and Lovell, C. A. K. (2001). Graph efficiency and productivity measures: An application to U.S agriculture. *Applied Economics*, 33(11):1433–1442.

Appendix A

Detailed Results for CT procedure

Table A.1: Illustrative Bank Branches Data

Unit	Staff Costs	Other Operating Costs	Deposit Accounts	Credit	Interest Revenue	FDH Eff.	VRS Eff.
B3	16.819	24.471	4892.629	10238.760	52.234		
B5	11.243	23.558	4777.107	8756.227	52.449		
B9	18.441	35.090	6450.385	12479.115	64.644		
B10	10.106	23.104	5223.611	12572.231	61.332	100%	100%
B11	15.129	32.781	7666.449	10221.426	67.682	100%	100%
B13	12.979	23.658	4991.984	10194.377	48.583		
B15	11.717	29.314	4070.630	6418.995	40.328		
B16	18.306	31.359	7561.477	21922.138	101.725	100%	100%
B17	16.505	31.574	6322.393	17323.595	81.404	100%	
B19	12.211	24.411	3663.067	10103.516	49.062		
B20	11.981	17.857	3899.831	10658.024	51.052	100%	100%
B21	12.689	25.489	4797.797	10281.063	48.822		
B22	16.166	26.062	3946.813	7358.401	46.214		
B26	12.041	19.688	5524.905	7393.716	48.912	100%	100%
B27	10.021	16.780	3394.509	8269.236	39.565	100%	100%
B29	12.739	18.505	5635.758	6667.397	63.048	100%	100%
B50	12.505	17.508	4745.698	9603.156	48.199	100%	100%
B51	15.178	21.418	5758.861	6007.936	64.210	100%	
B52	14.146	22.291	4391.541	8259.170	50.503	100%	
B53	12.959	20.117	5372.053	7323.490	64.076	100%	
B56	9.073	19.259	2888.434	8694.691	39.974	100%	100%
B57	9.747	13.004	2107.062	5012.420	24.202	100%	100%
B58	10.639	22.566	3344.774	10293.887	43.311	100%	
B59	13.338	24.820	4354.301	10889.840	57.033		
Average	13.195	23.529	4824.253	9872.617	54.523		
Max	18.441	35.090	7666.449	21922.138	101.725		
Min	9.073	13.004	2107.062	5012.420	24.202		
Stdev	2.646	5.494	1356.082	3640.361	15.456		

Table A.2: Results from Additive Units Invariant Model (values are rounded)

Unit	Peers	Slacks $(e_1, e_2, s_1, s_2, s_3)$	BRWZ Efficiency
B3	B16(0.503), B50(0.497)	(1.39, 0, 1268.6, 5557.1, 22.87)	68.30%
B5	B10(0.80), B16(0.113), B50(0.086)	(0, 0, 670.53, 4620.72, 12.33)	78.04%
B9	B16	(0.135, 3.73, 1111.09, 9443.02, 37.08)	64.70%
B13	B10(0.431), B16(0.247), B29(0.323)	(0, 0, 941.5, 2780.4, 23.3)	76.77%
B15	B10(0.804), B16(0.196)	(0, 4.6, 1612.31, 7990.23, 28.94)	53.58%
B17	B10(0.22), B16(0.78)	(0, 2.03, 725.72, 2545.43, 11.45)	85.36%
B19	B10(0.67), B16(0.22), B29(0.11)	(0, 0, 2122.7, 3870.92, 21.38)	68.42%
B21	B10(0.66), B16(0.31), B29(0.03)	(0, 0, 1152.3, 4970.8, 24.9)	71.42%
B22	B16(0.62), B50(0.38)	(0.078, 0, 2537.8, 9852.6, 35.04)	53.37%
B51	B16(0.26), B29(0.32), B50(0.42)	(1.09, 0, 0, 5859.3, 2.6)	79.29%
B52	B10(0.077), B16(0.314), B50(0.61)	(0, 0, 1276.16, 5444.8, 15.5)	71.41%
B53	B10(0.078), B16(0.11), B20(0.25), B29(0.55), B50(0.006)	(0, 0, 0, 2521.204, 0)	91.46%
B58	B10(0.83), B16(0.02), B29(0.15)	(0, 0, 1980.7, 1564.91, 18.98)	73.05%
B59	B10(0.462), B16(0.326), B29(0.212)	(0, 0, 1718.96, 3478.2, 17.84)	74.56%

Table A.3: Results from RAM Model (values are rounded)

Unit	Peers	Slacks $(e_1, e_2, s_1, s_2, s_3)$	BRWZ Efficiency
B3	B10(0.83), B16 (0.17)	(5.36, 0, 718.04, 3881.43, 15.79)	66.28%
B5	B10(0.945), B16 (0.055)	(0.69, 0, 575.05, 4330.11, 11.11)	77.14%
B9	B16	(0.13, 3.73, 1111.1, 9443.02, 37.08)	64.70%
B13	B10(0.933), B16(0.067)	(2.32, 0, 388.43, 3004.95, 15.5)	74.62%
B15	B10(0.804), B16(0.196)	(0, 4.6, 1612.31, 7990.23, 28.94)	53.58%
B17	B10(0.22), B16(0.78)	(0, 2.03, 725.72, 2545.43, 11.45)	85.36%
B19	B10(0.842), B16(0.158)	(0.81, 0, 1930.65, 3948.88, 18.66)	67.63%
B21	B10(0.71), B16(0.29)	(0.22, 0, 1101.04, 4991.63, 24.18)	71.23%
B22	B10(0.642)B16(0.358)	(3.12, 0, 2114.5, 8563.9, 29.6)	51.89%
B51	B10(0.285), B16(0.125), B29(0.59)	(2.5, 0, 0, 4243.7, 3.18)	77.68%
B52	B10(0.823), B29(0.177)	(3.6, 0, 904.9, 3269.15 11.13)	68.87%
B53	B10(0.25), B16(0.04), B29(0.72)	(0.66, 0, 234.62, 1365.72, 0)	91.00%
B58	B10(0.883), B29(0.117)	(0.225, 0, 1927.1, 1586.7, 18.22)	72.71%
B59	B10(0.79), B16(0.21)	(1.53, 0, 1355.25, 3625.82, 12.7)	73.24%

Table A.4: Results from CT Procedure (values are rounded)

Unit	Facet/Peers	$(h_1, h_2, g_1, g_2, g_3)$	BRWZ Efficiency
B3	F_6/B10(0.41), B11(0.26), B26(0.33)	(0.72, 1, 1.22, 1, 1.13)	77.50%
B5	F_7/B10(0.25), B26(0.36), B27(0.21), B29(0.17)	(1, 0.84, 1.05, 1, 1)	90.42%
B9	F_5/B11(0.81), B16(0.19)	(0.85, 0.93, 1.185, 1, 1.15)	80.52%
B13	F_1/B16(0.06), B29(0.07), B50(0.86)	(0.994, 0.78, 1, 1, 1.085)	86.42%
B15	F_2/B29(0.45), B50(0.14), B57(0.41)	(0.98, 0.55, 1, 1, 1.118)	73.83%
B17	F_5/B10(0.1), B11(0.32) B16(0.59)	(1, 0.98, 1.165, 1, 1.07)	92.27%
B19	F_2/B20(0.9), B29(0.02), B57(0.08)	(0.967, 0.715, 1.0337, 1, 1)	83.22%
B21	F_6/B10(0.37), B11(0.35), B26(0.28)	(0.98, 1, 1.28, 1, 1.23)	85.45%
B22	F_6/B11(0.49), B26(0.51)	(0.84, 1, 1.664, 1.2, 1.26)	68.49%
B51	F_5/B11(0.19), B16(0.01), B29(0.8)	(0.87, 1, 1.05, 1.24, 1)	86.03%
B52	F_6/B10(0.07), B11(0.18), B26(0.75)	(0.88, 1, 1.34, 1, 1.053)	84.52%
B53	F_5/B10(0.03), B11(0.09), B16(0.02), B29(0.86)	(1, 1, 1.086, 1.015, 1)	96.87%
B58	F_4/B10(0.41), B56(0.59)	(0.89, 0.924, 1.15, 1, 1.126)	83.45%
B59	F_1/B16(0.08), B20(0.76), B29(0.16)	(0.945, 0.767, 1.0278, 1, 1)	84.82%

Appendix B

Properties of the GDF defined in model (8.4)

G1 - $0 \leq G(\mathbf{x},\mathbf{y}) \leq 1$;

G2 - $G(\alpha\mathbf{x},\alpha^{-1}\mathbf{y}) \leq \frac{1}{\alpha^2}G(\mathbf{x},\mathbf{y}), \alpha \geq 1$ and $G(\alpha\mathbf{x},\alpha^{-1}\mathbf{y}) \geq \frac{1}{\alpha^2}G(\mathbf{x},\mathbf{y}), \alpha \leq 1$;

G3 - $G(\alpha\mathbf{x},\mathbf{y}) \ \leq \ \frac{1}{\alpha}G(\mathbf{x},\mathbf{y}) \leq G(\mathbf{x},\mathbf{y}), \alpha \geq 1$;

G4 - $G(\mathbf{x},\alpha\mathbf{y}) \ \leq \ \alpha G(\mathbf{x},\mathbf{y}) \leq G(\mathbf{x},\mathbf{y}), 0 \leq \alpha \leq 1$

G1 Proof: The GDF cannot be greater than 1. In order for this to happen the numerator in (8.4) should be greater than the denominator. However, as every θ_i in the numerator is ≤ 1, and every β_r in the denominator is ≥ 1, GDF > 1 results in an impossibility. This means that the maximum value of $G(\mathbf{x},\mathbf{y})$ is 1, happening when the numerator equals the denominator. As every θ_i in the numerator is ≤ 1, and every β_r in the denominator is ≥ 1, the equality between the numerator and denominator can only happen when all θ_i and all β_r are 1.

The GDF may be zero when some inputs (but not all, as we assume that it is not possible to produce outputs with zero inputs) are zero. For zero outputs the model cannot find a feasible solution as it would be possible to find an infinitely large β_{ro} associated with the zero output.

G2 Proof: This property states that $G(\mathbf{x},\mathbf{y})$, satisfies sub-homogeneity (see for example Russell, 1985) of degree minus two. Indeed,

$$G(\alpha\mathbf{x},\alpha^{-1}\mathbf{y}) = \min\left\{\frac{\alpha(\Pi_i\theta_i)^{1/m}}{\alpha^2\alpha^{-1}(\Pi_r\beta_r)^{1/r}} \mid (\theta_i(\alpha x_{io}),\beta_r(\alpha^{-1}y_{ro})) \in T\,,\ 0\le\theta_i\le 1,\ \beta_r\ge 1\right\}\Leftrightarrow$$
$$\min\left\{\frac{(\Pi_i(\alpha\theta_i))^{1/m}}{\alpha^2(\Pi_r(\alpha^{-1}\beta_r))^{1/r}} \mid ((\theta_i\alpha)x_{io},(\beta_r\alpha^{-1})y_{ro}) \in T,\ 0\le\theta_i\le 1,\ \beta_r\ge 1\right\}\Rightarrow$$
$$G(\alpha\mathbf{x},\alpha^{-1}\mathbf{y}) \le \frac{1}{\alpha^2}G(\mathbf{x},\mathbf{y}) \text{ for } \alpha\ge 1, \text{ and } G(\alpha\mathbf{x},\alpha^{-1}\mathbf{y}) \ge \frac{1}{\alpha^2}G(\mathbf{x},\mathbf{y}) \text{ for } \alpha\le 1$$

G3 and G4 Proof: These properties relate with the weak monotonicity properties of the geometric distance function. The input monotonicity implies that

$$G(\alpha\mathbf{x},\mathbf{y}) = \min\left\{\frac{\alpha\frac{1}{\alpha}(\Pi_i\theta_i)^{1/m}}{(\Pi_r\beta_r)^{1/r}} \mid (\theta_i(\alpha x_{io}),\beta_r y_{ro}) \in T\,,\ 0\le\theta_i\le 1,\ \beta_r\ge 1\right\}\Leftrightarrow$$
$$\min\left\{\frac{1}{\alpha}\frac{(\Pi_i(\alpha\theta_i))^{1/m}}{(\Pi_r(\beta_r))^{1/r}} \mid ((\theta_i\alpha)x_{io},\beta_r y_{ro}) \in T,\ 0\le\theta_i\le 1,\ \beta_r\ge 1\right\}\Rightarrow$$
$$G(\alpha\mathbf{x},\mathbf{y}) \le \frac{1}{\alpha}G(\mathbf{x},\mathbf{y}) \text{ for } \alpha\ge 1 \Rightarrow G(\alpha\mathbf{x},\mathbf{y}) \le G(\mathbf{x},\mathbf{y})$$

The output monotonicity implies that

$$G(\mathbf{x},\alpha\mathbf{y}) = \min\left\{\frac{(\Pi_i\theta_i)^{1/m}}{\alpha\frac{1}{\alpha}(\Pi_r\beta_r)^{1/r}} \mid (\theta_i x_{io},\beta_r(\alpha y_{ro})) \in T\,,\ 0\le\theta_i\le 1,\ \beta_r\ge 1\right\}\Leftrightarrow$$
$$\min\left\{\alpha\frac{(\Pi_i(\theta_i))^{1/m}}{(\Pi_r(\alpha\beta_r))^{1/r}} \mid (\theta_i x_{io},\beta_r(\alpha y_{ro})) \in T,\ 0\le\theta_i\le 1,\ \beta_r\ge 1\right\}\Rightarrow$$
$$G(\mathbf{x},\alpha\mathbf{y}) \le \alpha G(\mathbf{x},\mathbf{y}) \text{ for } \alpha\le 1 \Rightarrow G(\mathbf{x},\alpha\mathbf{y}) \le G(\mathbf{x},\mathbf{y})$$

Appendix C

Malquist Index based on the GDF

Input and Output Change Components

According to Färe et al. (1998, 2001) technical change is the product of input biased technical change (IBTECH), output biased technical change (OBTECH) and magnitude of technical change (MATECH), all defined in (C.1).

$$
\begin{aligned}
IBTECH &= \Big(\frac{D^{t+1}(x_t, y_t)}{D^t(x_t, y_t)} \times \frac{D^t(x_{t+1}, y_t)}{D^{t+1}(x_{t+1}, y_t)}\Big)^{\frac{1}{2}} \\
OBTECH &= \Big(\frac{D^t(x_{t+1}, y_{t+1})}{D^{t+1}(x_{t+1}, y_{t+1})} \times \frac{D^{t+1}(x_{t+1}, y_t)}{D^t(x_{t+1}, y_t)}\Big)^{\frac{1}{2}} \\
MATECH &= \frac{D^t(x_t, y_t)}{D^{t+1}(x_t, y_t)}
\end{aligned}
\tag{C.1}
$$

Consider that the distance functions in (C.1) are CRS input oriented and can be expressed as $D^t(x_t, y_t) = \frac{x_t^{*t}/x_t}{y_t^{*t}/y_t}$ for the single input/output case (note that Färe et al. (2001) defined the components in (C.1) for output oriented measures. As under CRS input and output efficiency measures are equal it does not matter the model orientation in calculating these components). We can, therefore, write the above (C.1) equivalently as (C.2).

$$
\begin{aligned}
IBTECH &= \Big(\frac{x_t^{*t+1}}{x_t^{*t}} \times \frac{x_{t+1}^{*t}}{x_{t+1}^{*t+1}}\Big)^{\frac{1}{2}} \\
OBTECH &= \Big(\frac{y_{t+1}^{*t+1}}{y_{t+1}^{*t}} \times \frac{y_t^{*t}}{y_t^{*t+1}}\Big)^{\frac{1}{2}} \\
MATECH &= \frac{x_t^{*t}/y_t^{*t}}{x_t^{*t+1}/y_t^{*t+1}}
\end{aligned}
\tag{C.2}
$$

Note that the IBTECH and the OBTECH are very similar to our input and output change components except that in each case one of the ratios is inverted. This inversion results from the existence of the magnitude of technical change component. In fact when multiplying all the 3 components in (C.2) we get (C.3), which is equivalent to our product of input change and output change components of technological change.

$$\left(\left(\frac{x_t^{*t+1}}{x_t^{*t}} \times \frac{x_{t+1}^{*t}}{x_{t+1}^{*t+1}}\right) \times \left(\frac{y_{t+1}^{*t+1}}{y_{t+1}^{*t}} \times \frac{y_t^{*t}}{y_t^{*t+1}}\right) \times \left(\frac{x_t^{*t}/y_t^{*t}}{x_t^{*t+1}/y_t^{*t+1}}\right)^2\right)^{\frac{1}{2}} \Leftrightarrow$$
$$\left(\left(\frac{x_{t+1}^{*t}}{x_{t+1}^{*t+1}} \times \frac{x_t^{*t}}{x_t^{*t+1}}\right) \times \left(\frac{y_{t+1}^{*t+1}}{y_{t+1}^{*t}} \times \frac{y_t^{*t+1}}{y_t^{*t}}\right)\right)^{\frac{1}{2}} \quad \text{(C.3)}$$

Equivalence with RD Approach

For the single input/output case the GDF and RD approaches are equivalent as long as efficiency models are the same in both procedures. If, for example, VRS radial input oriented targets are used to compute the GDF efficiency, then the equivalence between the technological change and the efficiency change components of the RD and GDF approaches can be readily seen in (9.5), since the GDF measure would equal the radial VRS input oriented measure. The equivalence between the residual component in (9.11) and the RD scale component is proved next for the single input/output case. When input oriented efficiency measures are calculated both in relation to a CRS (c) technology and in relation to a VRS (v) technology, then the scale component of the RD approach is defined by (C.4), where $\gamma_t^{t(r)} = x_t^{*t(r)}/x_t$, being $x_t^{*t(r)}$ the target input in frontier r that can be either c or v.

$$Scale_{RD} = \left(\frac{\gamma_{t+1}^{t(c)}/\gamma_{t+1}^{t(v)}}{\gamma_t^{t(c)}/\gamma_t^{t(v)}} \times \frac{\gamma_{t+1}^{t+1(c)}/\gamma_{t+1}^{t+1(v)}}{\gamma_t^{t+1(c)}/\gamma_t^{t+1(v)}}\right)^{1/2} = \left(\frac{x_{t+1}^{*t(c)}/x_{t+1}^{*t(v)}}{x_t^{*t(c)}/x_t^{*t(v)}} \times \frac{x_{t+1}^{*t+1(c)}/x_{t+1}^{*t+1(v)}}{x_t^{*t+1(c)}/x_t^{*t+1(v)}}\right)^{1/2} \quad \text{(C.4)}$$

Consider the CRS technology defined by $y = bx$, then we can replace projections on c by: $x_{t+1}^{*t(c)} = y_{t+1}/b_t$, $x_t^{*t(c)} = y_t/b_t$, $x_{t+1}^{*t+1(c)} = y_{t+1}/b_{t+1}$, and $x_t^{*t+1(c)} = y_t/b_{t+1}$, which results in (C.4) being equivalent to (C.5).

$$Scale_{RD} = \left(\frac{(y_{t+1}/b_t)/x_{t+1}^{*t(v)}}{(y_t/b_t)/x_t^{*t(v)}} \times \frac{(y_{t+1}/b_{t+1})/x_{t+1}^{*t+1(v)}}{(y_t/b_{t+1})/x_t^{*t+1(v)}}\right)^{1/2} = \left(\frac{\frac{x_t^{*t(v)}}{x_{t+1}^{*t(v)}}}{\frac{y_t}{y_{t+1}}} \times \frac{\frac{x_t^{*t+1(v)}}{x_{t+1}^{*t+1(v)}}}{\frac{y_t}{y_{t+1}}}\right)^{\frac{1}{2}} \quad \text{(C.5)}$$

The latter expression in (C.5) is exactly equivalent to the scale component in (9.11) since in input oriented measures target outputs are equal to observed outputs. The GDF residual component is, therefore, equal to the scale component of the RD approach, having the advantage of not requiring the computation of efficiency scores in relation to a CRS technology. Note that recently Orea (2002) developed a Malmquist index approach, in the parametric context, that also does not require the computation of efficiency scores in relation to CRS technologies. The GDF approach uses, instead of projections on the CRS frontier, output relationships on the VRS frontier to account for scale effects.

Appendix D

Transactional Efficiency Results

Table D.1: Descriptive Statistics of Transactional Data from January 2002 to September 2002

		Rent	N.ETMs	N.Clients	N.Registr	N.TrasCATs	N.Dep.ETMs
	Avg	3.136	2.538	3412.481	11.385	2920.38	892.846
Jan	Max	5.574	6	5812	42	23508	3607
2002	Min	0.816	1	971	1	0	136
	StDev	1.049	0.979	1097.692	7.814	3519.742	695.779
	Avg	3.136	2.558	3428.692	5.462	2397.68	764.231
Feb	Max	5.574	6	5833	13	21822	3577
2002	Min	0.816	1	995	0	0	83
	StDev	1.049	0.958	1098.76	3.427	3139.92	593.207
	Avg	3.136	2.558	3420.75	7.731	2624.4	863.885
Mar	Max	5.574	6	5823	26	22831	3592
2002	Min	0.816	1	1001	0	231	65
	StDev	1.049	0.958	1096.749	4.919	3337.967	664.822
	Avg	3.136	2.558	3429.885	9.635	2645.627	933.75
Apr	Max	5.574	6	5923	41	22390	3610
2002	Min	0.816	2	1022	1	0	101
	StDev	1.049	0.916	1102.347	8.698	3274.547	770.013
	Avg	3.136	2.558	3429.192	24	2702.25	998.577
May	Max	5.574	6	5928	84	24926	3874
2002	Min	0.816	2	1036	2	239	85
	StDev	1.049	0.916	1098.236	19.52	3550.697	788.563
	Avg	3.136	2.558	3165.731	41.308	2417.712	840.404
June	Max	5.574	6	6132	126	22991	2869
2002	Min	0.816	2	1002	3	199	60
	StDev	1.049	0.916	1100.724	30.657	3295.806	601.306
	Avg	3.136	2.558	3146.308	44.808	2869.288	969.519
July	Max	5.574	6	6049	207	25613	3496
2002	Min	0.816	2	1022	7	246	90
	StDev	1.049	0.916	1090.531	38.783	3686.977	718.145
	Avg	3.136	2.558	3140.019	23.615	2400.981	818.096
Aug	Max	5.574	6	5685	126	23263	3241
2002	Min	0.816	2	1051	0	284	54
	StDev	1.049	0.916	1065.9	21.498	3214.066	570.765
	Avg	3.136	2.558	3138.846	29	2451.385	832.962
Sep	Max	5.574	6	5661	66	23610	2817
2002	Min	0.816	2	1062	4	219	131
	StDev	1.049	0.916	1063.186	15.113	3289.966	572.978

Table D.2: Transactional Efficiency Scores for January 2002

Unit	VRS Eff.	GDF	Unit	VRS Eff.	GDF
B1	1	1	B32	1	1
B3	0.8943	0.5371	B33	1	1
B4	1	1	B34	0.7821	0.7764
B5	0.6345	0.6345	B35	0.4849	0.4357
B7	1	1	B36	1	1
B8	1	1	B37	0.5255	0.3566
B9	1	1	B38	0.5812	0.4969
B10	0.5860	0.5860	B39	0.8503	0.7290
B11	0.6439	0.6439	B40	0.9644	0.9318
B12	0.4975	0.4781	B41	1	1
B13	0.8551	0.8551	B42	1	1
B14	0.5421	0.4563	B44	1	1
B15	0.5846	0.4900	B45	1	1
B16	0.5169	0.5046	B46	1	1
B17	0.2447	0.2447	B48	1	1
B18	0.3015	0.2309	B49	0.5816	0.5056
B19	0.4297	0.4231	B50	0.5297	0.2729
B20	0.2211	0.2196	B51	0.5214	0.3903
B21	0.6439	0.6302	B52	0.4212	0.4053
B22	0.6234	0.5287	B53	0.2975	0.2462
B25	0.5115	0.4797	B54	1	1
B26	0.4127	0.3116	B55	0.2967	0.2209
B27	0.6655	0.4968	B56	0.4561	0.4367
B29	0.7404	0.7244	B58	0.4700	0.4144
B30	0.4228	0.3802	B59	0.4367	0.4367
B31	0.4931	0.4505	B60	1	1

Table D.3: Transactional Efficiency from January 2002 to September 2002

Unit	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Avg
B4	1	1	1	1	1	1	1	1	1	1
B7	1	1	1	1	1	1	1	1	1	1
B36	1	1	1	1	1	1	1	1	1	1
B46	1	1	1	1	1	1	1	1	1	1
B48	1	1	1	1	1	1	1	1	1	1
B54	1	1	1	1	1	1	1	1	1	1
B32	1	0.8671	1	1	1	1	1	1	1	0.9852
B44	1	1	0.6265	1	1	1	1	1	1	0.9585
B33	1	0.9295	1	1	1	1	1	0.8564	0.8231	0.9566
B8	1	1	0.5659	1	1	1	1	1	1	0.9518
B9	1	1	1	1	1	1	1	0.7756	0.5693	0.9272
B41	1	1	1	1	1	1	1	0.633	0.639	0.9191
B1	1	1	1	0.108	0.5107	1	1	1	1	0.84652
B3	0.5371	0.6968	1	1	0.7812	0.5051	0.8149	1	1	0.8150
B5	0.6345	0.828	0.6491	0.8517	0.7393	0.5521	1	0.9013	1	0.7951
B13	0.8552	0.8136	0.8655	0.7546	0.8777	0.7607	0.787	0.6921	0.7276	0.7927
B49	0.5056	1	0.7187	0.8767	1	0.6931	1	0.6166	0.5574	0.7742
B42	1	1	1	1	0.6735	0.4762	0.6927	0.4683	0.6046	0.7684
B21	0.6302	0.8505	0.5007	0.6793	0.7967	0.557	1	0.8022	0.8051	0.7357
B25	0.4797	0.8836	0.5786	0.4979	1	0.5953	1	0.8667	0.4712	0.7081
B22	0.5287	0.5385	0.7516	0.7963	0.6604	0.417	1	0.806	0.8534	0.7058
B45	1	1	0.5084	0.5524	0.4272	0.3892	0.4715	1	1	0.7054
B27	0.4968	1	0.3974	0.6242	0.6574	0.6853	0.8816	0.7509	0.7571	0.6945
B12	0.4781	0.5218	0.4284	0.561	0.8672	1	0.7804	0.8622	0.7426	0.6935
B11	0.6439	1	0.6495	0.5175	0.6345	0.4854	0.6038	0.6808	0.8344	0.6722
B38	0.4969	0.385	0.5603	0.8125	0.5357	0.5189	0.7016	0.8814	0.9736	0.6518
B34	0.7764	0.8601	0.6307	0.5875	0.7337	0.5616	0.4886	0.3954	0.7774	0.6457
B40	0.9318	0.807	0.6229	0.5957	0.6214	0.4358	0.3806	0.3144	1	0.6344
B16	0.5046	0.671	0.5846	0.5902	0.5033	0.6357	0.626	0.7086	0.7444	0.6187
B29	0.7244	0.6155	0.5325	0.7333	0.6061	0.4148	0.78	0.6508	0.4968	0.6171
B31	0.4505	0.6444	0.5635	1	0.8621	0.3941	0.4912	0.3686	0.5337	0.5898
B14	0.4563	0.6948	0.482	0.7435	0.5951	0.3663	1	0.4642	0.4286	0.5812
B52	0.4053	0.381	0.2959	0.2484	0.2277	1	1	0.4624	1	0.5579
B50	0.2729	0.4923	0.361	0.3684	0.4201	0.4863	1	1	0.5664	0.5519
B10	0.586	0.7196	0.549	0.4487	0.4081	0.445	0.4332	0.5209	0.6913	0.5335
B15	0.49	0.5648	0.4556	0.6606	0.656	0.3149	0.3808	0.5627	0.4503	0.5040
B39	0.729	0.5382	0.4112	0.6358	0.5463	0.5256	0.3228	0.2417	0.4449	0.4884
B56	0.4367	0.3556	0.5027	0.4697	0.4944	0.4235	0.6848	0.4103	0.6128	0.4878
B51	0.3903	0.386	0.5872	0.3859	0.3745	0.425	0.5451	0.5496	0.5031	0.4607
B59	0.4367	0.4765	0.2608	0.2901	0.2574	0.7261	0.606	0.4223	0.564	0.4489
B37	0.3566	0.459	0.2546	0.5881	0.404	0.4755	0.5424	0.5308	0.3632	0.4416
B26	0.3116	0.0103	0.5877	0.3265	0.3584	0.3416	0.6121	0.3418	1	0.4322

Transactional Efficiency Results

Unit	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Avg
B30	0.3802	0.6075	0.547	0.3109	0.3346	0.1898	0.4146	0.293	0.5825	0.4067
B60	1	0.0047	0.0021	0.0027	0.3452	0.4243	1	0.4836	0.345	0.4008
B35	0.4357	0.4925	0.3448	0.3222	0.2407	0.4787	0.2541	0.445	0.5744	0.3987
B19	0.4231	0.4373	0.7355	0.4505	0.1859	0.223	0.3865	0.2882	0.3032	0.3815
B58	0.4144	0.5678	0.2968	0.2972	0.2378	0.2206	0.4426	0.2815	0.5555	0.3683
B53	0.2462	0.1453	0.3693	0.3012	0.3055	0.2714	0.3089	0.2511	0.2638	0.2736
B20	0.2196	0.2712	0.3067	0.2809	0.2067	0.2718	0.2642	0.2341	0.3796	0.2705
B55	0.2209	0.2315	0.1418	0.4147	0.3717	0.2012	0.2125	0.1991	0.1675	0.2401
B17	0.2447	0.0076	0.2101	0.1721	0.521	0.2027	0.1838	0.2139	0.229	0.2205
B18	0.2309	0.229	0.1848	0.2028	0.1891	0.2564	0.1991	0.1932	0.2198	0.2117

Appendix E

Operational Efficiency Results

Table E.1: Descriptive Statistics of Operational Data from March 2001 to September 2002

		Rent	Staff	ΔCli	ΔCuracc	ΔOthre	ΔTdcp	ΔCredb	ΔCredas	Trans
	Avg.	2.896	5.093	-25.056	-143.867	78.179	-96.342	220.777	42.363	4062.41
Mar01	Max	5.347	7	33	120.949	1049.536	296.63	3708.493	349.398	7902
	Min	0.783	3	-192	-610.464	-2876.268	-303.344	-508.958	-241.982	8
	Avg.	2.911	5.161	-28.232	-17.941	196.103	-161.614	106.853	24.991	3999.44
Apr01	Max	5.347	7	341	582.99	1070.42	118.011	1330.548	250.85	7724
	Min	0.783	3	-1327	-534.497	-1091.13	-695.793	-263.924	-240.431	7
	Avg.	2.881	5.164	8.273	43.554	208.434	193.797	41.08	67.553	4399.37
May01	Max	5.347	7	234	514.969	2379.705	1993.78	583.499	2265.021	8421.28
	Min	0.783	3	-37	-454.609	-423.824	-2.13	-256.068	-280.06	1252.5
	Avg.	2.911	5.143	-5.929	4.104	-126.483	-285.695	82.917	17.998	4075.84
Jun01	Max	5.347	8	257	873.495	611.496	2590.851	592.417	255.973	7517.56
	Min	0.783	3	-1051	-910.875	-2623.039	-1102.762	-228.185	-103.805	1086.28
	Avg.	2.922	5.278	3.815	239.841	321.309	136.879	30.131	40.266	4743.61
Jul01	Max	5.347	8	36	1303.304	4996.618	981.48	299.438	264.722	9041.34
	Min	0.783	3	-78	-472.012	-704.103	-113.107	-735.912	-87.978	1383.14
	Avg.	2.945	5.291	4.527	257.941	10.572	-209.314	81.497	13.675	4644.11
Aug01	Max	5.347	8	252	999.337	584.152	19.478	1279.591	249.184	8087.39
	Min	0.783	3	-755	-891.033	-1227.247	-5196.202	-1175.986	-174.594	1545.96
	Avg.	2.885	5.37	3.796	-172.382	24.103	-161.971	13.791	8.5	3925.13
Sep01	Max	4.843	8	40	318.059	1558.06	81.658	681.583	248.591	7602.73
	Min	0.783	3	-112	-641.818	-640.222	-427.47	-370.841	-197.105	13
	Avg.	2.995	5.5	3.44	-72.259	231.255	86.282	70.562	40.236	4882.19
Oct01	Max	5.347	8	65	898.101	2255.23	752.357	609.94	636.082	8432.02
	Min	0.783	3	-104	-473.544	-381.106	-136.312	-300.685	-90.577	10
	Avg.	3.01	5.54	-196.5	0.193	79.749	-17.743	26.527	28.654	4597.39
Nov01	Max	5.347	9	2	814.547	465.977	684.096	761.839	339.447	7814.89
	Min	0.783	3	-628	-533.355	-291.732	-788.774	-897.113	-159.007	8.14
	Avg.	3.017	5.51	-5.98	316.848	233.677	-139.087	61.398	31.428	5621.37
Dec01	Max	5.347	9	168	1903.233	1196.881	-0.289	416.167	515.722	10009.5
	Min	0.783	3	-482	-298.775	-129.353	-2472.606	-303.464	-96.024	1879.62
	Avg.	3.147	5.471	40.275	141.586	75.227	-22.831	147.22	53.498	4419.08
Jan02	Max	5.574	8	155	845.9	636.9	200.2	758.6	321.7	7957

Operational Efficiency Results

		Rent	Staff	ΔCli	ΔCuracc	ΔOthre	ΔTdep	ΔCredb	ΔCredas	Trans
	Min	0.816	3	-74	-408.5	-1215.3	-217	-533.4	-251	1849
Feb02	Avg.	3.147	5.392	18.608	-116.098	-92	-127.725	23.176	-14.784	4615.39
	Max	5.574	7	66	509	438	167	591	336	7956
	Min	0.816	3	-65	-823	-2735	-327	-401	-1122	1762
Mar02	Avg.	3.147	5.392	-8.137	-50.706	135.451	127.333	63.451	22.392	3751.96
	Max	5.574	7	33	734	948	335	446	476	6747
	Min	0.816	3	-137	-918	-226	11	-263	-155	1227
Apr02	Avg.	3.147	5.412	8.941	-30.51	92.176	-58.157	48.627	5.078	4806.94
	Max	5.574	7	100	581	599	99	363	179	8687
	Min	0.816	3	-70	-409	-242	-172	-265	-205	1581
May02	Avg.	3.147	5.392	-1.039	156.451	-53.078	-83.333	51.882	8.941	5897.14
	Max	5.574	7	45	7441	364	1	401	172	10399
	Min	0.816	3	-127	-1643	-505	-288	-272	-153	1862
Jun02	Avg.	3.077	5.408	-331.49	-173.673	-84.061	-100.347	7.592	22.286	5232.35
	Max	5.574	7	-34	660	463	22	385	154	11422
	Min	0.816	3	-987	-7595	-393	-325	-310	-167	13
Jul02	Avg.	3.147	5.51	-19.51	246.745	-160.824	-140.02	10.373	13.686	6412.31
	Max	5.574	8	27	953	266	231	505	506	14133
	Min	0.816	3	-111	-1394	-2819	-460	-873	-247	11
Aug02	Avg.	3.147	5.627	-6.843	194.176	-20.098	-105.137	85.255	-44.216	5823.61
	Max	5.574	8	188	760	477	96	526	188	9701
	Min	0.816	3	-364	-560	-1573	-252	-304	-2056	2499
Sep02	Avg.	3.147	5.667	-0.98	-240.02	38.392	-156.588	80.784	-14.863	6342.63
	Max	5.574	8	29	362	418	-29	835	828	11185
	Min	0.816	3	-47	-760	-311	-431	-579	-509	2434

Table E.2: RDM Operational Efficiency values for April 2001

Unit	$1-\beta$	Norm	Unit	$1-\beta$	Norm
B1	1	1	B32	0.9602	0.925
B2	1	1	B33	1	1
B3	1	1	B34	1	1
B4	1	1	B35	0.8899	0.8002
B5	0.8388	0.7969	B36	0.9275	0.8601
B7	1	1	B37	0.8437	0.7859
B8	0.8878	0.8694	B38	1	1
B9	1	1	B39	0.726	0.704
B10	1	1	B40	0.8718	0.8405
B11	0.7274	0.6539	B41	1	1
B12	0.9061	0.8815	B42	0.7451	0.6456
B13	0.7984	0.7984	B43	0.8456	0.421
B14	1	1	B44	0.8673	0.8544
B15	0.8111	0.8	B45	0.9204	0.8748
B16	1	1	B46	0.889	0.8541
B17	0.7398	0.7319	B48	0.7041	0.7028
B18	0.7218	0.6989	B49	0.9282	0.8952
B19	0.7971	0.7878	B50	0.7941	0.7665
B20	0.8359	0.7608	B51	0.799	0.7778
B21	0.7785	0.7765	B52	0.7956	0.76
B22	0.8836	0.8526	B53	1	1
B23	1	1	B54	1	1
B25	0.9998	0.9165	B55	0.7632	0.7542
B26	0.9225	0.8537	B56	0.771	0.7388
B27	0.902	0.8276	B57	1	1
B28	1	1	B58	1	1
B29	0.9218	0.8485	B59	0.7841	0.5655
B30	0.8671	0.7622	B60	0.7211	0.7155
B31	0.9082	0.8404	Avg	0.891	0.854

Table E.3: Operational Efficiency Values from March 2001 to September 2002

	B2	B4	B9	B28	B23	B33	B54	B7	B27	B46	B1	B26
Mar	1	1	1	1	1	1	1	0.986	1	1	1	0.968
Apr	1	1	1	1	1	1	1	1	0.962	0.958	1	0.929
May	1	1	1	1	0.995	1	0.985	1	1	0.982	0.931	1
Jun	1	1	1	1	1	1	1	0.975	1	1	1	1
Jul	1	1	1	1	1	1	1	0.98	0.914	1	1	1
Aug	1	1	1	1	1	1	1	1	1	1	1	0.993
Sep	1	1	1	1	1	1	1	1	1	0.908	1	1
Oct		1	1			1	1	1	1	0.99	1	1
Nov		1	1			1	1	1	1	1	1	1
Dec		1	1			1	1	1	1	1	1	1
Jan		1	1			1	1	1	1	1	1	1
Feb		1	1			0.99	1	1	1	1	1	0.936
Mar		1	1			1	1	1	1	1	1	1
Apr		1	1			1	1	0.97	1	0.979	1	1
May		1	1			1	1	0.985	1	1	1	1
Jun		1	1			1	1	1	0.99	1	1	1
Jul		1	1			1	1	0.946	0.979	1	0.924	0.93
Aug		1	1			1	1	1	1	1	0.895	1
Sep		1	1			1	1	1	0.982	0.952	1	1
Avg	1	1	1	1	0.999	0.999	0.999	0.992	0.991	0.988	0.987	0.987

	B32	B49	B38	B29	B37	B51	B25	B12	B8	B34	B43	B5
Mar	1	1	1	1	1	0.907	1	0.983	0.939	1	1	0.965
Apr	0.974	0.968	1	0.971	0.93	0.914	1	0.962	0.943	1	0.914	0.936
May	0.729	0.945	1	0.944	0.916	0.986	1	0.869	0.856	1	0.934	0.84
Jun	1	0.965	1	0.989	0.997	0.972	1	0.998	1	1	1	0.977
Jul	1	0.906	1	1	0.991	0.957	1	0.986	0.925	1		0.945
Aug	1	0.996	1	1	1	1	1	1	0.957	1		1
Sep	1	1	0.906	1	0.873	1	0.962	1	0.968	1		1
Oct	1	0.995	0.984	0.983	0.992	1	1	1	0.998	1		1
Nov	1	1	1	0.943	1	1	0.932	1	0.922	1		1
Dec	1	1	1	1	1	1	0.951	0.969	1	1		0.987
Jan	1	1	0.862	1	1	1	0.808	1	1	0.708		1
Feb	1	1	1	1	0.939	1	1	1	0.988	0.952		0.864
Mar	1	1	1	0.963	1	1	1	1	0.988	0.899		1
Apr	1	1	0.99	0.969	1	1	0.964	0.955	1	1		1
May	1	1	1	0.966	0.957	0.881	0.905	1	1	1		0.902
Jun	1	1	1	0.955	1	1	0.958	0.902	0.917	1		0.893
Jul	0.986	0.881	0.9	1	0.908	0.88	1	0.881	0.963	1		0.98
Aug	0.977	1	1	1	1	1	1	1	1	0.82		0.976
Sep	1	1	1	0.929	1	1	1	0.93	1	1		1
Avg	0.982	0.982	0.981	0.979	0.974	0.974	0.973	0.970	0.967	0.967	0.962	0.961

	B16	B36	B44	B57	B14	B3	B39	B41	B17	B45	B10	B13
Mar	1	1	0.996	0.972	0.989	1	1	1	1	0.972	0.995	0.954
Apr	1	0.942	0.915	1	1	1	0.857	1	0.862	0.933	1	0.899
May	0.845	0.874	0.812	0.889	0.841	0.944	1	0.757	0.79	0.777	0.711	0.732
Jun	1	1	0.993	1	1	1	1	1	0.911	0.995	0.903	1
Jul	1	0.812	0.85	0.889	0.933	0.842	1	0.811	0.796	0.728	0.781	0.865
Aug	1	0.982	0.947	1	0.935	0.979	0.911	0.858	1	0.956	0.938	1

continues on next page

continued from previous page

Sep	1	0.95	0.942	0.927	0.849	1	0.872	1	1	0.926	0.891	0.945
Oct	1	0.937	0.986		0.973	1	0.911	0.936	0.977	0.928	1	0.982
Nov	1	1	0.988		1	1	1	1	0.876	1	1	0.917
Dec	1	0.992	0.941		0.976	1	1	0.89	0.886	0.989	0.979	1
Jan	1	0.758	1		0.929	0.968	0.734	0.996	1	0.759	1	1
Feb	1	1	1		0.958	0.882	0.737	1	0.908	1	0.946	1
Mar	1	0.95	1		1	0.942	1	1	1	0.997	0.967	1
Apr	1	1	1		1	0.998	0.984	1	1	1	1	1
May	0.619	1	0.904		0.89	0.763	1	0.916	0.944	1	0.939	0.73
Jun	1	1	1		0.9	0.871	1	0.826	0.967	0.954	0.853	1
Jul	0.967	1	0.903		0.857	0.827	1	0.915	1	1	0.859	0.944
Aug	1	0.961	1		1	0.958	0.841	0.904	0.949	0.934	1	0.925
Sep	0.746	1	0.981		0.863	0.872	0.958	1	0.924	0.945	1	0.872
Avg	0.957	0.956	0.956	0.954	0.942	0.939	0.937	0.937	0.936	0.936	0.935	0.935

	B58	B20	B59	B50	B60	B56	B35	B48	B31	B52	B53	B15
Mar	0.986	1	0.955	0.958	1	0.907	0.923	0.983	0.999	0.937	1	0.851
Apr	1	0.896	0.892	0.901	0.865	0.875	0.96	0.84	0.954	0.864	1	0.903
May	0.727	0.827	0.685	0.843	0.826	0.889	0.663	1	0.838	0.758	0.821	0.581
Jun	0.969	1	1	1	0.975	0.959	0.969	1	0.952	0.949	0.993	0.957
Jul	0.795	0.958	0.778	0.802	0.716	0.89	0.64	0.732	0.798	0.743	0.679	0.642
Aug	0.94	0.963	0.977	0.978	0.996	0.929	0.962	0.981	0.944	0.98	0.949	1
Sep	1	1	1	1	1	0.995	1	0.99	0.965	0.97	0.905	1
Oct	1	0.966	0.942	0.969	1	0.937	0.962	0.997	0.952	0.949	0.951	1
Nov	1	1	1	0.918	1	0.885	0.951	0.991	0.909	1	0.85	0.891
Dec	1	0.759	1	0.966	0.896	0.939	1	0.973	0.963	0.914	0.921	1
Jan	0.61	0.868	1	0.833	0.721	1	0.745	0.595	0.653	0.689	1	1
Feb	0.934	1	0.94	0.94	0.87	0.961	1	1	0.912	0.942	0.852	0.754
Mar	1	1	0.883	1	0.932	0.867	1	0.957	1	0.987	1	1
Apr	1	0.964	0.993	0.944	0.96	0.969	1	1	0.926	1	1	0.946
May	1	0.935	1	0.783	0.979	0.912	1	0.868	0.764	0.828	0.895	1
Jun	0.942	1	0.89	0.942	0.961	0.891	1	0.972	1	0.983	0.887	1
Jul	0.898	0.715	0.78	0.889	0.963	0.866	0.86	1	1	0.998	0.816	1
Aug	0.949	0.878	1	0.984	1	0.933	0.921	0.684	0.942	1	1	0.952
Sept	0.991	0.981	0.954	1	0.992	0.971	1	1	0.935	0.898	0.861	0.896
Avg	0.934	0.932	0.929	0.929	0.929	0.925	0.924	0.924	0.916	0.915	0.915	0.914

	B18	B19	B42	B21	B30	B22	B40	B55	B11
Mar	1	0.935	1	0.969	1	0.931	1	1	0.919
Apr	0.856	0.888	0.879	0.885	0.895	0.951	0.95	0.871	0.858
May	0.784	1	0.786	0.714	0.659	0.773	0.712	0.702	0.603
Jun	0.955	0.939	0.988	0.991	0.964	0.971	1	0.951	0.96
Jul	0.769	0.72	0.683	0.633	0.678	0.615	1	0.654	0.689
Aug	0.935	0.886	0.977	0.897	1	0.909	0.724	0.917	0.945
Sept	0.923	1	1	1	1	1	0.888	0.903	0.861
Oct	0.919	0.958	0.955	0.996	0.933	0.96	0.917	0.941	0.906
Nov	0.862	0.903	1	0.909	0.941	0.923	1	0.868	0.856
Dec	1	0.92	0.946	0.717	0.957	1	0.952	0.907	0.899
Jan	1	0.81	1	0.698	0.686	0.805	0.757	0.611	0.517
Feb	0.916	1	0.837	1	0.886	0.949	1	0.833	1
Mar	1	0.887	0.798	0.954	0.99	0.939	0.917	0.806	0.996
Apr	0.93	0.947	0.948	0.908	1	0.948	0.968	0.93	0.928

continues on next page

continued from previous page									
May	0.778	1	0.82	0.967	0.921	0.701	0.684	0.928	0.756
Jun	0.884	0.812	1	1	0.849	1	0.922	0.897	1
Jul	0.822	0.77	0.785	1	0.962	0.687	0.826	0.98	0.994
Aug	1	0.881	0.873	0.9	0.862	1	0.866	0.928	1
Sep	0.944	0.958	0.902	1	0.957	0.874	0.809	0.901	0.773
Avg	0.909	0.906	0.904	0.902	0.902	0.891	0.889	0.869	0.866

Appendix F

Profit Efficiency Results

Table F.1: Descriptive Statistics of Profit Data from March 2001 to September 2002

Month		staff	supplycost	curracc	othress	credb	credass
Mar01	Avg.	5.175	18.678	4021.641	10992.344	9208.759	823.665
	Stdev	1.212	5.02	1699.14	5818.065	3953.25	578.268
Apr01	Avg.	5.175	18.567	4001.255	11190.634	9313.16	848.054
	Stdev	1.167	4.798	1711.72	589.55	4022.515	588.862
May01	Avg.	5.175	20.075	4059.819	11433.269	9370.595	923.353
	Stdev	1.167	5.192	1706.895	5837.918	4006.987	721.739
Jun01	Avg.	5.158	20.67	4062.163	11310.599	9451.313	943.256
	Stdev	1.177	5.271	1683.258	5804.599	4042.199	743.193
Jul01	Avg.	5.304	21.404	4376.328	11835.790	9634.139	1002.412
	Stdev	1.264	5.746	1723.450	5916.581	3962.608	772.968
Aug01	Avg.	5.304	18.998	4634.860	11852.928	9718.892	1016.392
	Stdev	1.264	4.972	1805.170	5860.543	3996.889	805.414
Sep01	Avg.	5.393	22.014	4498.158	11936.475	9836.372	1032.304
	Stdev	1.275	5.89	1757.222	5839.758	4118.799	818.153
Oct01	Avg.	5.50	23.529	4641.283	12722.801	10234.985	1087.140
	Stdev	1.094	7.126	1678.760	5894.368	4137.364	835.661
Nov01	Avg.	5.519	57.269	4651.288	12823.965	10273.746	1117.136
	Stdev	1.163	14.113	1703.675	5932.287	4100.099	875.577
Dec01	Avg.	5.519	42.939	4963.146	13055.727	10338.148	1148.915
	Stdev	1.094	11.374	1861.745	5988.665	4110.672	924.294
Jan02	Avg.	5.48	40.12	5106.85	13128.62	10487.02	1201.65
	Stdev	1.06	9.82	1926.82	6082.02	4126.58	932.97
Feb02	Avg.	5.404	45.69	4990	13041.13	10510.04	1187.35
	Stdev	1.05	9.5	1915.72	6065.79	4099.78	930.22
Mar02	Avg.	5.404	40.3	4935	13175	10573.79	1210.13
	Stdev	1.05	8.35	1911.28	6121.59	4062.76	958.15
~~Apr02~~	Avg.	5.423	45.689	4905.442	13262.615	10624.327	1213.038

Month		staff	supplycost	curracc	othress	credb	credass
	Stdev	0.997	10.362	1933.798	6156.369	4054.567	973.78
May02	Avg.	5.404	45.716	5060.538	13212.558	10671.885	1222.423
	Stdev	0.995	9.843	2232.69	6162.727	4036.728	970.845
Jun02	Avg.	5.481	42.037	4988.538	13355.769	10955.327	1257.865
	Stdev	1.057	9.85	1867.059	6211.758	4246.039	975.537
Jul02	Avg.	5.519	50.92	5237.038	13198.788	10973.019	1271.808
	Stdev	1.075	12.63	1899.967	6176.087	4208.483	990.269
Aug02	Avg.	5.635	49.085	5439.365	13179.75	11061.5	1228.615
	Stdev	1.048	11.538	1944.554	6124.707	4235.697	850.433
Sep02	Avg.	5.673	44.632	5200.462	13218.115	11149.365	1217.596
	Stdev	1.061	9.571	1837.424	6122.08	4206.47	841.139

Table F.2: Long Run Profit Efficiency Results for April 2001

Unit	Tech. GDF	Facet	Actual Profit	Tech. Profit	Max. Profit	Profit GDF	Alloc. GDF	Scale Effect	Mix Effect	α	γ
B1	0.533	F14	106.32	138.13	678.45	0.3095	0.5806	0.7494	0.7748	2.632	3.512
B2	1	F3,F5...	97.31	97.31	675	0.3643	0.3643	0.7759	0.4694	2.89	3.724
B3	0.794	F1	378.22	451.78	673.94	0.6119	0.7711	1.0971	0.7029	1.451	1.323
B4	0.878	F3	273.06	286.66	679.64	0.6434	0.7325	0.9394	0.7797	1.952	2.078
B5	0.865	F2	361.25	365.99	677.17	0.6485	0.7501	1.0441	0.7184	1.624	1.556
B7	0.839	F12	239.25	259.08	674.34	0.6273	0.7476	0.8886	0.8413	2.102	2.366
B8	0.818	F3	151.63	157.71	677.45	0.3768	0.4604	0.6814	0.6757	2.496	3.663
B9	0.755	F8	439.4	616.62	677.42	0.5575	0.7381	No	0.7381	1.07	0.793
B10	1	F1	466.3	466.3	680.99	0.7068	0.7068	1.1042	0.6401	1.4	1.268
B11	1	F6	527.28	527.28	676.51	0.5962	0.5962	1.1368	0.5245	1.167	1.026
B12	0.799	F13	534.19	557.8	680.57	0.7422	0.9286	No	0.9286	1.333	0.998
B13	0.857	F2	368.78	391.45	675.15	0.6354	0.7414	1.0899	0.6803	1.503	1.379
B14	1	F13	518.18	518.18	676.51	0.8784	0.8784	1.0876	0.8077	1.211	1.114
B15	0.580	F6	271.65	511.47	680.47	0.4561	0.7858	1.1543	0.6807	1.282	1.110
B16	1	F1,F4...	676.33	676.33	676.33	1	1	1	1	1	1
B17	1	F8	663.91	663.91	677.01	0.6776	0.6776	No	0.6776	1.017	0.682
B18	0.832	F3	353.62	359.15	674.54	0.6597	0.7925	1.267	0.6255	1.845	1.456
B19	0.610	F1	274.99	475.63	677.86	0.4555	0.7462	1.0931	0.6826	1.4	1.281
B20	0.817	F12	355.96	377.56	680.27	0.7046	0.8628	1.1377	0.7583	1.777	1.562
B21	0.785	F1	357.19	456.85	678.47	0.5903	0.752	1.1005	0.6833	1.437	1.305
B22	0.676	F3	255.98	268.47	675.22	0.4236	0.6265	0.9291	0.6743	2.016	2.17
B23	0.604	F3	118.04	144.26	677.95	0.3613	0.5986	0.7865	0.7611	2.588	3.291
B25	0.803	F5	158.06	170.27	678.20	0.5643	0.7025	1.1601	0.6055	2.456	2.117
B26	0.694	F2	346.91	412.99	676.22	0.5278	0.761	1.2295	0.619	1.642	1.335
B27	0.799	F12	309.18	319.86	675.9	0.6679	0.8362	1.1634	0.7187	1.946	1.673
B28	1	F1,F2...	344.87	344.87	680.09	1.2023	1.2023	No	1.2023	2.333	0.822
B29	1	F1,F2...	520.27	520.27	676.03	0.8571	0.8571	No	0.8571	1.614	0.891
B30	0.760	F10	213.12	252.84	677.45	0.4769	0.6273	1.1778	0.5326	1.865	1.583
B31	0.785	F1	431.51	491.8	676.11	0.606	0.7719	1.0854	0.7111	1.328	1.223
B32	0.801	F1	361.24	467.32	676.77	0.6282	0.7846	1.0976	0.7148	1.434	1.306
B33	1	F2,F3	309.51	309.51	677.48	0.6101	0.6101	1.0313	0.5916	1.84	1.784
B34	1	F1,F2...	479.38	479.38	674.09	0.9407	0.9407	No	0.9407	1.4	0.825
B35	0.795	F8	360.88	415.1	678.27	0.681	0.8569	1.1667	0.7344	1.424	1.220
B36	0.3791	F3	40.13	90.25	675.13	0.1199	0.3163	0.5281	0.5989	2.964	5.613
B37	1	F2,F3...	340.83	340.83	678.52	0.8496	0.8496	1.1952	0.7108	1.807	1.512
B38	1	F2,F4	430.21	430.21	678.76	0.769	0.769	1.1473	0.6702	1.4	1.22
B39	0.736	F1	354.06	460.73	673.19	0.5202	0.7068	1.1042	0.6401	1.4	1.268
B40	0.835	F4	389.24	425.42	678.53	0.7185	0.8603	1.0401	0.8271	1.456	1.400
B41	0.786	F4	486.33	535.21	676.41	0.7272	0.9248	1.0115	0.9143	1.187	1.174
B42	0.644	F1	334.41	458.36	677.05	0.4749	0.7378	1.1017	0.6696	1.424	1.293
B43	1	F14	28.71	28.71	678.97	0.1497	0.1497	0.3495	0.4282	3.381	9.671
B44	0.874	F3	239.61	242.97	678.01	0.5337	0.6109	0.9604	0.6362	2.206	2.297
B45	0.546	F14	110.19	145.99	677.69	0.2819	0.5161	0.7867	0.6561	2.540	3.229
B46	0.922	F1	487.72	490.27	676.61	0.7618	0.826	1.0862	0.7605	1.372	1.263
B48	0.735	F6	381.02	522.26	679.31	0.5921	0.8056	1.1508	0.7	1.25	1.086

Profit Efficiency Results

Unit	Tech. GDF	Facet	Actual Profit	Tech. Profit	Max. Profit	Profit GDF	Alloc. GDF	Scale Effect	Mix Effect	α	γ
B49	0.894	F7	494.99	518.26	675.89	0.8108	0.9066	1.0892	0.8324	1.31	1.202
B50	0.943	F2	422.28	434.65	677.25	0.8303	0.8802	1.308	0.6729	1.685	1.288
B51	0.713	F6	484.41	524.40	679.33	0.551	0.7733	No	0.7733	1.333	0.956
B52	0.834	F2	391.15	449.23	677.11	0.6717	0.805	1.2585	0.6396	1.55	1.232
B53	1	F1,F5...	550.37	550.37	677.99	0.836	0.836	No	0.836	1.499	0.758
B54	1	F3,F5...	57.8	57.8	675.99	0.2062	0.2062	0.4302	0.4793	3.262	7.581
B55	0.835	F12	269.49	311.24	677.95	0.6777	0.8117	1.008	0.8052	1.925	1.91
B56	0.606	F1	285.02	484.91	679.49	0.4671	0.7713	1.1776	0.655	1.4	1.189
B57	0.903	F3	177.65	179.84	676.83	0.5325	0.5894	0.7985	0.7381	2.481	3.107
B58	0.751	F2	290.28	337.29	678.79	0.5868	0.7815	1.0173	0.7682	1.809	1.779
B59	0.887	F1	395.82	457.87	676.18	0.7505	0.8459	1.1348	0.7454	1.489	1.312
B60	1	F7,F8...	369.93	369.93	677.76	0.7239	0.7239	1.3325	0.5433	1.47	1.103
Avg	0.8299		340.97	382.44	677.31	0.6128	0.7332	1.0243	0.71497		

Table F.3: Short Run Profit Efficiency Results for April 2001

Unit	Tech. GDF	Actual Profit	Tech. Profit	Max. Profit	Profit GDF	Alloc. GDF	Scale Effect	Mix Effect	α	γ
B1	0.9275	106.3184	108.4292	143.0929	0.7989	0.8613	1.0436	0.8253	1.0436	1
B2	1	97.3051	97.3051	129.7447	0.8648	0.8648	0.8651	0.9997	1.1246	1.3
B3	0.7154	378.2209	442.4014	504.6245	0.7481	1.0458	1.2419	0.842	1.2419	1
B4	0.8386	273.0558	310.0788	361.0262	0.8045	0.9594	1.1351	0.8452	1.2086	1.0647
B5	0.7992	361.2471	454.3743	478.1158	0.8052	1.0076	1.0515	0.9582	1.0515	1
B7	0.8676	239.2481	270.0231	318.8085	0.7728	0.8906	1.0969	0.812	1.0969	1
B8	0.699	151.634	206.528	207.0281	0.6797	0.9724	1.0109	0.9619	1.0109	1
B9	0.7049	439.3986	454.5746	586.3657	0.7392	1.0487	1.3077	0.8019	1.3077	1
B10	1	466.3009	466.3009	592.5124	1.0111	1.0111	1.0452	0.9674	1.3	1.2438
B11	1	527.2822	527.2822	609.4086	0.8253	0.8253	1.1734	0.7034	1.1667	0.9943
B12	0.8015	534.1887	557.5434	624.7948	0.9995	1.2471	No	1.2471	1.3003	0.7687
B13	0.8115	368.7802	452.0504	486.7443	0.839	1.0339	1.1067	0.9342	1.1067	1
B14	1	518.1845	518.1845	622.2477	0.9719	0.9719	1.0876	0.8936	1.2111	1.1135
B15	0.5481	271.6529	377.2726	377.911	0.5385	0.9824	1	0.9824	1	1
B16	1	676.3324	676.3324	676.3324	1	1	1	1	1	1
B17	1	663.9061	663.9061	665.0342	0.9331	0.9331	No	0.9331	1.0014	0.9726
B18	0.7514	353.6209	470.381	470.5757	0.7449	0.9914	1.0283	0.9641	1.0283	1
B19	0.7075	274.994	370.4218	373.803	0.6189	0.8747	No	0.8747	0.9458	1
B20	0.8565	355.9633	406.9885	468.6585	0.8195	0.9568	1.1863	0.8066	1.1863	1
B21	0.7752	357.1945	441.9217	474.4561	0.7933	1.0233	1.0904	0.9386	1.0904	1
B22	0.617	255.9806	351.5182	352.5352	0.5928	0.9608	No	0.9608	0.9941	1
B23	0.9226	118.0366	120.7283	158.0336	0.8132	0.8814	1.0956	0.8045	1.0956	1
B25	0.87	158.058	177.7347	212.3759	0.7516	0.8638	1.0151	0.851	1.0151	1
B26	0.7925	346.9126	415.3316	463.1748	0.7122	0.8986	No	0.8986	0.9758	1
B27	0.9201	309.1799	319.8178	408.2276	0.8152	0.886	1.1177	0.7927	1.1177	1
B28	1	344.8654	344.8655	422.9926	1.0534	1.0534	No	1.0534	1.3	0.9197
B29	1	520.266	520.266	582.671	1.1787	1.1787	No	1.1787	1.3	0.7607
B30	0.6777	213.118	290.4398	291.7372	0.6465	0.954	No	0.954	0.961	1
B31	0.7855	431.5106	477.4531	569.5499	0.8441	1.0746	1.257	0.8549	1.257	1
B32	0.8332	361.2376	428.5133	478.6704	0.7938	0.9527	1.0478	0.9093	1.0478	1
B33	1	309.5077	309.5077	404.8235	0.9202	0.9202	0.9529	0.9657	1.2387	1.3
B34	1	479.3802	479.3802	547.4322	1.0965	1.0965	No	1.0965	1.3	0.8689
B35	0.7268	360.8813	443.9229	481.8726	0.7093	0.976	1.0563	0.9239	1.1322	1.0718
B36	0.8174	40.1334	53.4502	71.2375	0.5385	0.6588	No	0.6588	0.7053	1
B37	1	340.8308	340.8308	434.4143	0.977	0.977	1.0825	0.9026	1.3	1.201
B38	1	430.2069	430.2069	532.6527	1.0213	1.0213	1.0881	0.9386	1.3	1.1947
B39	0.6935	354.0585	401.8385	474.5472	0.7338	1.0582	1.2665	0.8355	1.2665	1
B40	0.7899	389.2439	463.279	515.0694	0.7797	0.9871	1.2079	0.8172	1.2079	1
B41	0.8482	486.3283	534.627	602.3233	0.861	1.0152	1.1795	0.8607	1.1795	1
B42	0.6363	334.4092	382.0119	453.031	0.6566	1.0319	1.265	0.8157	1.265	1
B43	1	28.7062	28.7062	44.5861	0.6914	0.6914	0.8286	0.8344	1.0771	1.3
B44	0.7863	239.611	309.3946	319.979	0.7248	0.9218	1.0409	0.8856	1.0409	1

Profit Efficiency Results

Unit	Tech. GDF	Actual Profit	Tech. Profit	Max. Profit	Profit GDF	Alloc. GDF	Scale Effect	Mix Effect	α	γ
B45	0.7516	110.1883	125.4642	163.6107	0.5385	0.7164	No	0.7164	0.8722	1
B46	0.8747	487.7246	502.341	629.0793	0.9446	1.0799	1.3545	0.7973	1.3545	1
B48	0.6529	381.022	460.8869	512.3651	0.6883	1.0542	1.2047	0.875	1.2047	1
B49	0.9166	494.9954	501.6312	647.3476	0.9186	1.0022	1.2775	0.7845	1.2775	1
B50	0.9193	422.2828	445.0494	552.0547	0.9132	0.9934	1.206	0.8237	1.2712	1.0541
B51	0.7927	484.4114	500.085	597.7329	0.904	1.1403	1.3865	0.8224	1.3865	1
B52	0.7983	391.1534	492.4345	517.6991	0.7794	0.9764	1.0214	0.9559	1.0214	1
B53	1	550.3749	550.3749	596.9433	1.2371	1.2371	No	1.2371	1.3	0.7248
B54	1	57.7969	57.7969	79.5834	0.7992	0.7992	0.8214	0.9729	1.0678	1.3
B55	0.8264	269.4907	320.3126	357.584	0.7954	0.9624	1.1567	0.832	1.1567	1
B56	0.7342	285.0171	381.6092	384.3587	0.6546	0.8916	No	0.8916	0.9612	1
B57	0.7964	177.6528	221.7154	238.8651	0.7175	0.901	1.0595	0.8504	1.1324	1.0688
B58	0.7659	290.2821	386.4964	388.4463	0.6999	0.9139	1.0176	0.8981	1.0176	1
B59	0.8289	395.8153	474.0849	522.5592	0.8176	0.9864	1.0778	0.9152	1.0778	1
B60	1	369.932	369.932	457.0951	0.9695	0.9695	1.2196	0.7949	1.3	1.066
Avg	0.8452	340.9725	380.4621	432.2552	0.8175	0.9682	1.1085	0.8992		

Table F.4: Profit Efficiency Values from March 2001 to September 2002

	B2	B16	B17	B28	B53	B60	B37	B12	B29	B49	B50	B33
Mar	1	1	1	1	1	1	1	1	1	0.804	0.867	1
Apr	1	1	1	1	1	1	1	0.799	1	0.855	0.919	1
May	1	1	1	1	1	1	1	1	1	0.849	1	1
Jun	1	1	1	1	1	1	1	1	1	1	1	1
Jul	1	1	1	1	1	1	1	1	1	1	0.883	1
Aug	1	1	1	1	1	1	1	1	1	1	1	0.839
Sep	1	1	1	1	1	1	0.863	1	1	1	1	1
Oct		1	1		1	1	1	1	0.854	1	0.962	0.72
Nov		1	1		1	1	1	1	0.845	1	0.778	0.75
Dec		1	1		1	1	1	1	1	0.865	0.913	0.686
Jan		1	1		1	1	1	1	1	1	0.89	0.759
Feb		1	1		1	1	1	1	0.888	1	0.882	0.885
Mar		1	1		1	1	1	1	0.876	1	0.993	1
Apr		1	1		1	1	1	0.865	1	1	0.813	1
May		1	1		1	1	1	0.8	1	1	0.865	1
Jun		1	1		1	1	1	1	1	1	0.865	1
Jul		1	1		1	1	1	1	1	1	1	0.857
Aug		1	1		1	1	1	1	1	0.81	0.966	0.909
Sep		1	1		1	1	1	1	1	0.876	0.941	1
Avg	1	1	1	1	1	1	0.993	0.972	0.972	0.95	0.923	0.916

	B34	B14	B26	B7	B20	B46	B38	B27	B54	B4	B10	B18
Mar	1	1	0.552	0.569	0.667	0.823	0.834	0.611	0.306	0.701	1	0.704
Apr	1	1	0.547	0.638	0.771	0.826	1	0.652	1	0.839	1	0.745
May	1	1	0.621	0.592	0.745	0.79	1	0.607	0.373	0.775	1	0.707
Jun	1	1	0.677	0.619	1	0.883	1	0.796	0.424	0.664	0.769	0.652
Jul	1	1	0.67	0.637	1	0.853	1	0.636	1	0.682	0.82	0.679
Aug	1	1	0.751	0.717	1	0.872	1	0.624	0.402	0.716	1	0.727
Sep	1	1	0.801	0.805	0.844	1	0.824	0.668	0.418	0.733	0.715	0.668
Oct	1	0.745	1	1	0.867	0.923	0.824	1	1	0.692	1	0.601
Nov	1	0.874	1	1	0.786	0.936	0.783	0.906	0.634	0.686	0.682	1
Dec	0.669	0.793	1	1	0.862	0.797	0.677	1	1	0.72	0.686	0.828
Jan	1	0.829	1	1	0.771	0.762	0.678	1	1	1	0.616	1
Feb	1	0.865	1	1	0.765	0.776	0.652	1	1	0.8	0.674	1
Mar	1	0.811	1	1	0.768	0.884	0.738	1	1	0.863	0.638	1
Apr	0.67	0.657	1	1	1	0.858	0.687	1	1	1	0.645	0.757
May	0.647	0.791	1	1	0.916	0.767	1	1	1	1	0.668	0.768
Jun	1	0.819	1	1	0.877	0.775	0.91	0.845	1	1	0.626	0.835
Jul	0.688	0.822	1	1	0.825	1	0.779	0.774	1	1	0.765	0.729
Aug	1	0.852	1	1	0.972	0.757	1	0.797	1	0.753	1	0.755
Sep	0.689	0.895	1	1	0.931	0.962	0.788	0.783	1	0.872	0.747	0.802
Avg	0.914	0.882	0.875	0.872	0.861	0.855	0.851	0.826	0.819	0.816	0.792	0.787

	B35	B11	B1	B32	B48	B23	B25	B40	B13	B39	B41	B3
Mar	0.803	1	0.34	0.562	0.718	0.463	0.477	0.845	0.8	0.567	0.741	0.627
Apr	0.729	1	0.44	0.618	0.658	0.552	0.582	0.782	0.775	0.59	0.767	0.679

continues on next page

continued from previous page

May	0.707	0.575	0.253	0.575	1	1	0.719	0.8	0.632	0.624	0.723	0.609
Jun	0.737	1	0.543	0.789	1	1	1	0.831	0.78	0.679	0.817	0.65
Jul	0.706	1	1	0.741	1	1	1	0.812	0.791	0.682	0.805	0.556
Aug	0.745	0.587	1	0.901	1	0.757	1	0.804	0.823	0.72	0.812	0.599
Sep	0.809	1	0.355	1	1	0.589	0.507	0.77	0.597	0.73	0.814	0.719
Oct	0.795	1	1	0.793	1		0.841	0.767	0.752	0.64	1	0.775
Nov	0.741	0.711	1	0.756	1		0.731	0.752	0.65	0.74	0.724	0.78
Dec	0.697	0.606	1	0.668	0.474		0.81	0.657	0.715	0.733	0.695	0.841
Jan	0.684	0.579	1	0.674	0.817		0.668	0.734	0.752	0.756	0.621	0.763
Feb	0.892	0.603	1	0.748	0.672		1	0.719	0.807	0.808	0.671	0.828
Mar	0.831	0.572	1	0.652	0.612		1	0.799	0.873	0.974	0.628	0.876
Apr	0.772	0.665	1	0.731	0.601		0.7	0.626	0.754	0.866	0.68	0.819
May	0.751	0.676	1	1	0.589		0.685	0.67	0.709	0.873	0.806	0.859
Jun	0.817	0.537	1	1	0.757		0.7	0.706	0.844	0.727	0.774	0.843
Jul	0.878	1	1	1	0.59		0.714	0.812	0.875	0.806	0.66	0.781
Aug	0.829	1	0.478	0.696	0.579		0.694	0.815	0.676	0.926	0.736	0.661
Sep	1	0.791	0.492	0.75	0.561		0.667	0.778	0.825	0.866	0.728	0.762
Avg	0.785	0.784	0.784	0.771	0.77	0.766	0.763	0.762	0.759	0.753	0.748	0.738

	B51	B52	B59	B31	B5	B58	B57	B19	B55	B30	B44	B9
Mar	0.643	0.568	0.634	0.661	0.879	0.554	0.606	0.365	0.63	0.547	0.534	0.588
Apr	0.611	0.691	0.759	0.661	0.775	0.61	0.796	0.455	0.765	0.546	0.599	0.595
May	0.631	0.605	0.639	0.685	0.752	0.503	0.647	0.518	0.606	0.522	0.429	0.585
Jun	0.651	0.728	0.869	0.801	0.767	0.71	0.679	0.739	0.684	0.592	0.61	0.69
Jul	0.634	0.762	0.841	0.757	0.755	0.679	0.731	0.725	0.743	0.564	0.649	0.617
Aug	0.673	0.781	0.853	0.771	0.751	0.653	0.742	0.727	0.7	0.613	0.644	0.663
Sep	0.715	0.7	0.645	0.756	0.636	0.518	0.568	0.551	0.65	0.487	0.412	0.645
Oct	0.722	0.699	0.829	0.727	0.851	1		0.79	0.72	0.618	0.826	0.693
Nov	0.685	0.732	0.614	0.713	0.571	0.769		0.631	0.637	0.464	0.621	0.686
Dec	0.709	0.728	0.705	0.737	0.591	0.642		0.702	0.709	0.648	0.766	0.593
Jan	0.8	0.675	0.747	0.684	0.57	0.651		0.734	0.599	0.603	0.748	0.587
Feb	0.837	0.822	0.697	0.75	0.65	0.659		0.707	0.673	1	0.784	0.746
Mar	0.823	0.708	0.585	0.687	0.692	0.648		0.771	0.675	1	0.751	0.729
Apr	0.82	0.699	0.731	0.661	0.543	0.835		0.816	0.683	1	0.691	0.527
May	0.853	0.683	0.734	0.681	0.6	0.765		0.687	0.689	0.705	0.729	0.589
Jun	0.922	0.727	0.709	0.651	0.633	0.727		0.694	0.636	0.653	0.656	0.539
Jul	0.801	0.714	0.683	0.672	0.69	0.726		0.745	0.778	0.617	0.679	1
Aug	0.702	1	0.753	0.694	0.675	0.669		0.775	0.62	0.604	0.88	0.571
Sep	0.748	0.78	0.763	0.742	0.65	0.659		0.788	0.702	1	0.745	1
Avg	0.736	0.726	0.726	0.71	0.686	0.683	0.681	0.68	0.679	0.673	0.671	0.665

	B43	B8	B15	B21	B56	B22	B42	B45	B36
Mar	0.207	0.579	0.426	0.776	0.455	0.511	0.587	0.319	0.169
Apr	1	0.538	0.475	0.56	0.489	0.407	0.526	0.309	0.138
May	0.345	0.467	0.493	0.613	0.488	0.487	0.518	0.309	0.147
Jun	1	0.44	0.573	0.716	0.632	0.538	0.532	0.33	0.17
Jul		0.411	0.538	0.643	0.63	0.553	0.532	0.342	0.148
Aug		0.41	0.538	0.678	0.658	0.535	0.525	0.331	0.166
Sep		0.499	0.59	0.56	0.452	0.456	0.55	0.332	0.141

continues on next page

continued from previous page

Oct		0.645	1	0.741	0.586	0.705	0.589	0.296	0.173
Nov		0.422	1	0.562	0.486	0.601	0.594	0.318	0.164
Dec		0.638	0.603	0.507	0.427	0.599	0.546	0.299	0.164
Jan		0.87	0.545	0.495	0.6	0.615	0.464	0.263	0.148
Feb		0.691	0.504	0.518	0.592	0.45	0.475	0.32	0.155
Mar		0.672	0.507	0.501	0.749	0.455	0.471	0.305	0.162
Apr		0.792	0.528	0.537	0.591	0.5	0.478	0.384	0.141
May		0.781	0.523	0.537	0.632	0.515	0.548	0.368	0.151
Jun		0.667	0.521	0.44	0.701	0.613	0.558	0.439	0.218
Jul		1	0.59	0.541	0.513	0.616	0.581	0.444	0.18
Aug		0.753	0.624	0.51	0.709	0.627	0.62	0.577	0.213
Sep		0.694	0.577	0.613	0.523	0.641	0.597	0.339	0.189
Avg	0.638	0.63	0.587	0.581	0.574	0.549	0.542	0.349	0.165

Appendix G

Characteristics of Some Units

Table G.1: Characteristics of Star Branches in Profit Efficiency Assessment

Unit	staff	supplycost	curacc	othrcss	credb	credass	Eff.
B16	7.263	49.612	8066.043	21357.171	22055.931	4081.433	1
B17	7.158	45.14	6765.995	30635.916	17271.131	512.821	1
B53	5.316	32.698	5265.835	27683.379	7838.712	939.305	1
B60	5.895	35.02	4639.721	10110.47	20020.396	1517.45	1
B37	5	27.64	4412.904	7350.558	14337.5	1644.2	0.993
B12	5.947	33.209	7001.419	20713.546	8198.233	1628.081	0.972
B29	5.684	30.03	5501.041	24750.743	7590.218	918.958	0.972
B49	6	35.501	6129.963	16031.855	17352.602	1291.793	0.951
B50	5	30.298	5171.234	16231.155	9686.86	955.648	0.923
B33	4.632	26.627	4542.486	11146.462	6507.857	954.578	0.916
B34	6.579	42.833	6518.171	11484.254	10038.058	3833.909	0.914
Avg.	5.861	35.328	5819.528	17954.137	12808.863	1661.652	

Table G.2: Characteristics of Star Branches in Operational Efficiency Assessment

Unit	Rent	Staff	ΔCli	ΔCuracc	ΔOthre	ΔTdep	ΔCredb	ΔCredas	Trans	Eff.
B16	4.116	7.263	-38.316	65.885	125.153	-136.353	66.656	163.842	7367.462	0.957
B17	3.282	7.167	-76.667	97.267	99.453	-100.19	11.573	5.928	5613.392	0.936
B53	2.85	5.316	-49.053	41.067	152.474	-97.264	121.268	25.687	4252.911	0.915
B60	2.564	5.889	-46.944	59.882	188.747	-120.74	111.465	48.236	4566.898	0.929
B37	1.968	5	-23.421	36.935	-7.655	-66.886	32.215	11.252	5174.434	0.974
B12	2.886	5.947	-38	84.501	78.399	-79.435	-32.712	43.323	5534.339	0.97
B29	1.995	5.684	-58.895	31.971	205.762	-10.355	107.934	1.51	5313.347	0.98
B49	2.168	6	-17.526	47.56	-99.455	-92.761	140.9	17.19	6347.687	0.982
B50	2.432	5	-20.789	38.764	67.832	-138.924	32.085	3.5	4349.868	0.929
B33	1.411	4.632	-10.895	26.411	117.756	-44.145	141.275	52.042	4956.13	0.999
B34	3.742	6.579	-29.211	34.918	5.611	-40.639	78.225	65.364	6952.102	0.967
Avg.	2.674	5.862	-37.247	51.378	84.916	-84.336	73.717	39.807	5493.506	0.958

Table G.3: Characteristics of 'high operational low profit' Branches in Profit Efficiency Assessment

Unit	staff	supplycost	curacc	othrcss	credb	credass	Eff.
B1	3.158	22.452	1233.367	4246.511	6125.156	141.63	0.784
B3	5.842	36.662	5146.415	14349.904	10415.017	960.337	0.738
B4	4.947	25.625	3497.183	6895.901	9884.464	987.336	0.816
B5	5.789	37.06	5087.909	11589.564	9245.241	1160.073	0.686
B7	4.053	23.914	2828.456	8945.375	9565.342	543.145	0.872
B8	4.053	23.938	2545.001	5327.535	5832.126	338.082	0.630
B9	6.842	46.918	6963.326	14689.54	13199.921	849.329	0.665
B10	5.579	44.342	6799.294	14475.95	13312.284	933.233	0.792
B13	5.105	40.456	5132.718	11479.138	11554.592	1005.516	0.759
B14	5.947	39.35	6927.643	15155.203	11999.245	1678.782	0.882
B15	5.789	44.379	4484.678	9156.73	7528.869	1988.944	0.587
B18	4.895	26.88	3643.793	14097.046	5469.409	889.799	0.787
B19	5	35.87	3516.093	12338.948	10434.321	871.681	0.680
B20	4.842	29.721	4074.331	14067.01	10577.617	842.31	0.861
B21	5.895	43.76	5282.476	9514.855	10988.465	912.16	0.581
B25	3.579	27.206	2406.556	6200.732	5517.369	886.174	0.763
B26	5.053	30.955	7310.997	17118.322	7803.521	470.384	0.875
B27	4.211	28.044	3607.45	12337.987	8560.907	604.757	0.826
B30	4.789	34.904	3041.87	5772.76	11043.731	756.171	0.673
B31	6.632	39.132	5861.794	11938.399	14054.838	1011.08	0.710
B32	5.579	45.014	5944.356	12913.75	15074.366	1102.286	0.771
B35	5.421	32.707	4801.012	10999.704	8664.396	1652.478	0.785
B36	5.421	31.653	1343.262	1598.525	2115.445	260.551	0.165
B38	5.474	39.428	5865.756	10129.897	13621.833	1514.047	0.851
B39	5.789	34.011	5713.158	13896.801	9378.405	796.205	0.753
B41	7	45.534	6875.555	13880.874	18243.991	1457.897	0.748
B42	6.526	41.125	5261.374	8622.247	10075.7	853.19	0.542
B44	4	29.32	3258.651	8555.808	6245.581	521.851	0.671
B45	5.842	34.322	2173.584	2731.72	6373.257	770.588	0.349
B46	6	36.936	6600.129	17090.376	13536.133	1060.318	0.855
B48	7.368	48.405	5767.661	10098.512	10807.286	3829.272	0.770
B51	6.158	35.726	5635.401	22245.933	6740.77	672.407	0.736
B52	5.368	33.615	4427.868	17213.87	8893.71	745.648	0.726
B54	3	19.1	987.269	3708.058	2588.288	117.863	0.819
B56	4.842	30.985	2863.764	11396.459	8843.323	363.629	0.574
B58	4.947	32.411	3663.692	10910.757	10557.64	683.094	0.683
B59	5.474	38.402	4586.996	17352.417	11156.681	855.192	0.726
Avg.	5.303	34.872	4463.806	11163.328	9622.412	948.309	0.716

Table G.4: Characteristics of 'high operational low profit' Branches in Operational Efficiency Assessment

Unit	Rent	Staff	ΔCli	ΔCuracc	ΔOthre	ΔTdep	ΔCredb	ΔCredas	Trans	Eff.
B1	3.364	3.158	-6.316	16.14	33.148	-31.074	37.295	6.031	2208.064	0.987
B3	2.281	5.842	-47.263	20.942	174.814	-47.768	3.797	30.228	5419.024	0.939
B4	0.799	4.947	-33.526	31.76	44.445	-34.571	58.316	11.722	5140.391	1
B5	2.952	5.789	-27.789	20.08	96.643	-56.379	75.916	40.661	6100.028	0.961
B7	1.7	4.053	-18.105	27.512	19.723	-70.358	65.396	-6.2	3389.231	0.992
B8	2.406	4.053	-7.105	27.418	74.976	-29.167	113.797	10.591	3711.011	0.967
B9	2.875	6.842	-25.263	99.056	160.749	-84.239	106.579	21.512	8731.141	1
B10	4.188	5.611	-77.333	20.231	-25.317	-115.45	40.389	20.372	6002.553	0.935
B13	3.133	5.105	-16.895	23.779	87.37	-98.457	208.732	31.122	5616.958	0.935
B14	2.794	5.947	-38.947	3.199	41.174	-73.155	-18.437	-62.676	6832.741	0.942
B15	4.501	5.789	-16	56.005	180.283	-39.137	134.564	19.768	3817.275	0.914
B18	2.615	4.895	-31.053	-1.904	-46.505	-69.759	-67.995	5.784	3354.116	0.909
B19	4.681	5	-40.722	6.464	-47.838	-86.054	39.865	8	4649.342	0.906
B20	2.941	4.842	-37.474	40.344	89.271	-38.469	191.397	11.881	4172.708	0.932
B21	4.529	5.833	-14.556	-5.092	29.9	-186.654	54.323	17.363	5217.659	0.902
B25	3.136	3.579	9.263	80.401	69.385	-16.709	94.039	19.748	2929.668	0.973
B26	1.71	5.053	-19	207.284	203.697	-106.091	64.405	3.34	3728.942	0.987
B27	1.516	4.211	-28.158	15.942	12.843	-74.137	44.795	5.774	3296.415	0.991
B30	4.801	4.789	-8.474	28.813	30.382	-45.631	83.355	6.04	3895.175	0.902
B31	3.208	6.632	-42.684	29.407	-28.693	-77.047	24.338	45.996	6305.545	0.916
B32	5.461	5.556	-39.389	5.527	-90.445	-65.596	0.444	38.043	7244.984	0.982
B35	3.786	5.421	-22.842	10.685	15.948	-42.599	21.473	6.979	6018.097	0.924
B36	2.031	5.421	-14.368	24.603	53.739	-9.93	79.034	15.401	2277.241	0.956
B38	3.155	5.474	-18.684	11.536	15.649	-44.444	183.138	51.123	6412.353	0.981
B39	3.521	5.722	-17.389	-0.521	304.764	-127.265	27.434	23.708	4788.392	0.937
B41	4.876	7	-21.789	34.367	212.683	-93.849	206.572	14.445	6013.574	0.937
B42	3.279	6.526	-20.316	-28.473	-9.574	-103.187	-15.471	23.147	5931.557	0.904
B44	3.232	4	-3.105	41.138	-38.458	-34.786	45.86	16.147	3536.108	0.956
B45	2.673	5.842	0.211	14.766	58.254	-20.051	110.067	3.307	3035.953	0.936
B46	1.902	6	-23.947	39.005	159.709	-64.747	100.603	-5.339	5879.374	0.988
B48	3.852	7.368	-16.895	23.683	104.317	-81.094	-1.431	94.381	6005.281	0.924
B51	1.868	6.158	-29.579	25.665	-18.306	-72.438	103.795	14.3	5981.46	0.974
B52	3.043	5.368	-26.211	58.916	92.237	-89.991	116.502	9.359	4521.186	0.915
B54	2.092	3	19	31.452	124.557	-2.87	84.437	6.119	1316.045	0.999
B56	2.194	4.842	-42.368	5.102	-45.453	-75.587	8.032	1.835	3278.388	0.925
B58	3.421	4.947	-10.789	27.015	158.409	-73.302	135.89	20.02	4643.46	0.934
B59	3.802	5.474	5.211	27.709	179.95	-95.841	29.368	8.612	4596.842	0.93
Avg.	3.083	5.293	-21.852	29.719	67.053	-66.847	69.96	15.898	4750.097	0.948