International Series in
Operations Research And Management Science

Peter Bogetoft
Lars Otto

Benchmarking with DEA, SFA, and R

Springer

International Series in Operations Research & Management Science

Volume 157

For further volumes:
http://www.springer.com/series/6161

Peter Bogetoft • Lars Otto

Benchmarking with DEA, SFA, and R

Springer

Peter Bogetoft
Department of Economics
Copenhagen Business School CBS
Porcelaenshaven 16 A
2000 Frederiksberg
Denmark
pb.eco@cbs.dk

Lars Otto
Institute of Food and Resource
Economics
University of Copenhagen
Rolighedsvej 25
1958 Frederiksberg C
Denmark
lo@foi.dk

ISSN 0884-8289
ISBN 978-1-4419-7960-5 e-ISBN 978-1-4419-7961-2
DOI 10.1007/978-1-4419-7961-2
Springer New York Dordrecht Heidelberg London

To

Victoria, Rasmus, Stina, and Nete

Mathias and Gerda

Preface

Subject

This book covers recent advances in efficiency evaluation, most notably the Data Envelopment Analysis (DEA) and Stochastic Frontier Analysis (SFA) methods. It introduces the underlying theories, shows how to make the relevant calculations and discusses some applications. The aim is to make the reader aware of the pros and cons of the different methods and to train him or her on the proper usage of these methods in both standard and non-standard cases.

Several software packages have been developed that can be used to solve some of the most common DEA and SFA models. In this book, however, we rely on R, a free software environment that can be used for optimization, statistical computing, and graphics. This program enables the reader to solve not only standard problems but also many other problem variants. Using R, one can focus on understanding the business case and developing a good model. One is not restricted to predefined models or to the use of a one-size-fits-all approach.

There are several R routines that support the use of DEA and SFA models. While writing this book, we have also developed an R-package named *Benchmarking* that makes applications easy without limiting the variations in the models and calculations that innovative researchers and practitioners seek to use.

Audience and style

The intended audience includes graduate students, advanced consultants and practitioners with an interest in quantitative performance evaluation.

This book uses mathematical formulations of models, assumptions, etc. Unlike original contributions on this subject, however, this book de-emphasizes formal proofs, partially by placing them in appendices or by referring to the original

sources. Moreover, this book emphasizes the use of theories and interpretations of the mathematical formulations.

A series of small examples and graphical illustrations will be presented. This text also combines formal models with less formal economic and organizational thinking. Moreover, it discusses numerous applications based on projects on which we have worked. This includes some large projects with significant practical effects: e.g., the design of benchmarking-based regulations for energy companies in different European countries or the development of merger decision support systems for competition authorities.

Acknowledgements

The book covers material that we have used in several courses. We gratefully acknowledge comments by students at Copenhagen University and Copenhagen Business School, Denmark; Universitat Autònoma de Barcelona, Spain; the University of New England, UNE, Australia; the Helsinki School of Economics and Business Administration, Finland; and the Yale School of Management, USA.

In addition, we have benefitted from questions raised by participants in a series of industry courses that we presented in Denmark, Sweden, Germany, the Netherlands and Canada during the last five years. We would also like to acknowledge the influence of a large number of friends, coauthors and colleagues who have helped us to understand the basics of efficiency evaluation. At the risk of leaving out important individuals, we would like to acknowledge the contributions of Professors Per Agrell, Mette Asmild, Rajiv Banker, Robert Chambers, Tim Coelli, William W. Cooper, Philippe Van den Eeckaut, Rolf Färe, Finn Førsund, Shawna Grosskopf, Harold Fried, Emili Griffle-Tatje, Arne Henningsen, Pekka Korhonen, Jens Leth Hougaard, Knox Lovell, Ole Olsen, Niels Chr. Petersen, Leopold Simar, Jørgen Tind, Henry Tulkens, and Paul Wilson.

Copenhagen,
September 2010

Peter Bogetoft
Lars Otto

Contents

Acronyms and Symbols

List of main abbreviations used in the text

DEA	Data Envelopment Analysis
SFA	Stochastic Frontier Analysis
KPI	Key Performance Indicators
ADD	Additive technology
CRS	Constant Returns to Scale
DRS	Decreasing Returns to Scale, formally non-increasing returns to scale (NIRS)
FDH	Free Disposability Hull
FRH	Free Replicability Hull
IRS	Increasing Returns to Scale, formally non-decreasing returns to scale (NDRS)
VRS	Variable Returns to Scale
RTS	Returns To Scale

Lists of standard notation and symbols used in the mathematical formulations

T	Technology set
K	Number of firms, decision making units
k	Counter for firms
x	Input, often a vector
y	Output, often a vector
m	Number of inputs
n	Number of outputs
i	Counter for inputs
j	Counter for outputs
w	Input prices, often a vector
p	Output prices, often a vector
E	Input efficiency
F	Output efficiency
G	Graph efficiency

$\mathbb{R}$	Set of real numbers
$\mathbb{R}_+$	Set of non-negative real numbers
$\mathbb{R}_{++}$	Set of positive real numbers
EV	Expected value
exp	Exponential function

Chapter 1
Introduction to Benchmarking

This chapter gives a overview of the questions and methods covered in this book. What is the idea of efficiency analyses and benchmarking? Why do we benchmark? What are the state–of–the–art methods that we shall discuss in this book?

The chapter is mainly conceptual and can be read with a minimum of technical skills. The idea is to get an elementary and intuitive introduction to the subject. Still, to get a flavor of the approach of the book, we do formalize a few simple production economic models.

We first discuss why to benchmark, including learning, coordination and motivation. We then sketch the economist's ideal of a performance evaluation and the practical problems of using this in real contexts. We also discuss a common practical approach of using one or a few Key Performance Indicators (KPIs). We explain the fallacy of such approaches and the need for more model based, systematic benchmarking where a technology is estimated and performance is measures hereto. Lastly, we provide a brief introduction to the main methods covered in this book, Data Envelopment Analysis (DEA) and Stochastic Frontier Analyses (SFA)

1.1 Why benchmark

Relative performance evaluations or—using modern terminology—benchmarking is the systematic comparison of the performance of one firm against other firms.

More generally, it is comparison of production entities. The idea is that we compare entities that transform the same type of resources to the same type of products and services. The production entities can be firms, organizations, divisions, industries, projects, decision making units, or individuals. For convenience, we talk simply about the comparison of firms.

Benchmarking can be used in many different settings. It can be used to make *intra-organizational* comparisons, as when a headquarters wants to promote costs efficiency in its different subunits. Motivating a combination of profit and service objectives in a chain of fast food outlets is an obvious example; the owners can

P. Bogetoft and L. Otto, *Benchmarking with DEA, SFA, and R*, International Series in Operations Research & Management Science 157, DOI 10.1007/978-1-4419-7961-2_1,

evaluate the individual managers by comparing the sales and cost measures of such outlets. The owners can formalize the evaluations and introduce performance based payment schemes to motivate appropriate behavior.

Benchmarking can also be—and most often is—used to make *inter-organizational* comparisons . A primary example that we shall often refer to involves a regulator seeking to induce cost–efficiency or to avoid the misuse of monopoly power among a set of firms enjoying natural monopoly rights in different regions.

Lastly, benchmarking can be used to make *longitudinal* , panel, or dynamic comparisons, where the performance of one or more firms in different time periods are compared. Such comparisons are of considerable interests to economists and politicians since the development of productivity is an important driver of welfare improvements.

It is worthwhile emphasizing that the use of benchmarking is not restricted to *for-profit* organizations. Modern benchmarking methods can handle multiple objectives that are not explicitly aggregated. This opens the door for usage in *non-profit* / organizations, including most public organizations where there is no single objective or success criterion like profit maximization. Indeed, the ability to handle multiple objectives is one explanation of the popularity and numerous applications of modern benchmarking techniques.

In more general terms, the objectives of benchmarking can be related to one or more of the basic issues in any economic system, namely learning, coordination and motivation. Or using accounting terminology, benchmarking can be used to facilitate decision making (learning and coordination) and control (motivation). Although the preliminaries of performance assessment exercises normally contain arguments from all three categories, the design and execution of the model often reveals the importance associated to each task.

1.1.1 Learning

The stated objective of most benchmarking studies is to learn or get insight per see.

This is certainly the case in *scientific studies* where researchers examine the relative efficiency of firms in an industry, the relative efficiency of one industry against another, or the impact of some policy measure on industry performance.

Often, this is also the stated objective in *industry applications*. When several firms compare their performance, the official objective is often to support the learning and efficiency improvement of individuals. Firms are interested to know how well they are doing compared to others and which ones they can learn from. The nonparametric (Data Envelopment Analysis DEA) approaches that we shall cover in this book provide particular strengths in such cases as the peers or the dominating firms provide valuable and concrete information for performance improvement targets. Moreover, the various decompositions of the overall efficiency can point towards more specific means to improve efficiency, e.g. to change the scale of operation or the mix of resources used if scale or allocative efficiency is low. Still, the

actual operational changes will necessitate in-depth process benchmarking that may, or may not, be promoted by the participating firms. Competition may for obvious reasons limit the sharing of information about best practices.

Recent advances in *interactive benchmarking* is an attempt to push the learning perspective by allowing individual firms in an benchmarking exercise to define the comparison basis (potential peers), the objective (e.g. cost reduction or sales expansion), the aspiration level (e.g. to be in the top-ten) etc. of the evaluations. It has typically been used in industries where firms sees themselves as colleagues more than competitors, e.g. among waterworks, energy–networks, and farmers.

Practical application: Danish Waterworks

In Denmark, the industry organization Danish Water and Waste Water Association DANVA has for several years worked with benchmarking. In the early years, they relied on a traditional Key Performance Indicators, cf. below. Later they started undertaking more advanced benchmarking using Data Envelopment Analyses and made yearly reports for the sector in general and for the individual water works. In 2007, they took a further step towards active use of data and benchmarking to support learning of best practices. They introduced an interactive benchmarking system IBEN. The system enables individual waterworks to make multi-dimensional performance evaluations that reflect their own focus, conditions, mission and aspiration level. A series of pre-specified models cover both the totality of activities and significant individual processes and supports both economic and technical efficiency analyses, including energy-efficiency. Thus for example, individual managers can choose which processes to focus at, which other waterworks to compare to and which particular improvement strategies to explore, e.g. a cost reducing road to efficiency or a service expansion strategy. Similar applications have also been developed to support individual learning in several other industries, including the energy sector, the financial sector, and the health sector.

1.1.2 Coordination

In some studies, the objectives of the benchmarking explicitly addresses the allocation of tasks and possibly the restructuring of firms or the industry. Such studies may facilitate *coordination*, i.e. ensuring that the right firms are producing the right products at the right time and place. Coordination lies at the heart of much of traditional microeconomic theory and management science.

In firms and industries, benchmarks, tournaments and bidding schemes are used extensively to coordinate operations at optimal cost and performance. The headquarters of a bank for example may benchmark operations, not only to motivate local managers, but also to allocate resources and staff according to their profile.

Some coordination support requires noting more than ordinary benchmarking models where the performances of different entities are evaluated. This holds for example for tournaments. The use of the models in this book will allow these tournaments to be more effective by relying on more comprehensive performance analyses.

Other coordination support requires the use of more advanced benchmarking models to evaluate the *structural efficiency* of a set of entities. This may necessitate calculations in *networks* of individual benchmarking models. The methods covered in this book have been used for example, to evaluate the structural efficiency of whole industries and the possible gains from reallocating production factors and production rights across hundreds of production entities. They have been used also to decompose aggregate inefficiency into inefficiencies in the production units with given resources and the misallocation of resources among the units. They can be used to estimate gains from reallocating pollution right, production right etc, and to evaluate the possible gains from mergers as wee shall see in later chapters.

A interesting finding in such studies is that a better coordination may be just as valuable as the learning of best practice. This is relevant since it may be optimistic to suppose that all firms can adopt best practices. Also, for economic scholars, this insight is interesting since the economic tool box contains many mechanisms to facilitate better allocation, most notably the establishment of some sort of market, and much less hard theory on the internal processes of firms and organizations, and methods to boost the learning process.

Practical application: Reallocation of agricultural production

Estimates of structural efficiency have been used in Danish agriculture to inform the restructuring of *sugar beet production* following a change in EU subsidies. The new sugar regime led to dramatic price reductions: 36 per cent for white sugar and 40 percent for sugar beets over a four-year period starting in 2006. In collaboration with the monopsonist sugar beet processor, Danisco, and the association of sugar beet producers, we investigated the gains from reallocating production between the farmers. This involved the development of en sector model based on a series of individual benchmarking models and the calculation of possible cost reductions from reallocating production to take advantage of efficiency differences as well as economies of scale and scope. As a result of the study, it was concluded that sugar beet production could continue (although at a reduced scale) if appropriate measures was taken in terms of reducing processing capacity and reallocating primary production. One of three factories was closed, and a new sugar beet exchange was established in which more than 1200 farmers traded production rights to better allocate primary production. This exchange has since been repeated annually to ensure structural efficiency in a dynamic environment.

1.1.3 Motivation

A last general application of benchmarking is to facilitate incentive provision. By establishing the performance of an employee, a manager or a firm more precisely, it is possible to better target the incentives.

There are as usual several aspects of this . One is the *pre-contractual asymmetric information* or *adverse selection problem* of making it possible for better informed agents to extract information rent by claiming too high costs. Another is the *post-contractual moral hazard problem* arising from the inability of a principal to precise monitor if an agent pursues private objectives and perhaps shirks. Benchmarking can limit both of these incentive problems . Adverse selection can be limited by extracting information about an agent's type from past behavior. Moral hazard can be limited by relative performance evaluations, i.e. by announcing ex ante that performance based payments in the coming period will depend on the outcome of a benchmarking study to be done ex post.

The relationship between the benchmarking model and the motivational aspect may be implicit or explicit. An *implicit* or informal relationships is found when the mere existence of a model improves behavior because the performance now gets more attention in the minds of the agents. A more *explicit* and formalized relationship is found when the model is used to calculate individual inefficiencies and common measures of technological progress that are incorporated in the general control system. One can for example tie the budgeting rules, the salary plans or the tariff regulations directly to the outcome of the benchmarking.

To illustrate the different ways a benchmarking exercise may link up with the incentives, we may consider the regulation of electricity distribution in different countries. We shall return to this case in several chapters and give a more extended treatment in Chap. 10.

Practical application: Regulation of Electricity Networks in Europe

Electricity distribution is a natural monopoly industry with different firms serving different concession areas. This means that any given consumer (household, firm) can only buy the necessary distribution services from one provider, often referred to as a DSO (distribution system operator). This may lead to excessive costs and/or profits as well as to sub–optimal quality, since the DSO are not subject to a competitive pressure to cut costs, lower prices and improve quality. In most countries, a regulator appointed by the state is therefore allowed to interfere in the operations and in particular in the tariffs these companies charge. Unfortunately, the regulator lacks information about minimal costs. The asymmetric information can however be undermined using benchmarking as part of the regulation.

Sweden in several years relied on *light handed* regulation where the regulator only monitors performance and interferes on occasion. The development of a model—like the DEA models developed since 2000—signals a commitment of the

regulator to undermine the informational asymmetry and to keep up the pressure from regulatory oversight despite of increased complexity in the industry.

In Norway, the regulator has long committed to a more *heavy handed* mechanic approach relating allowed revenue to measures of general technological progress and individual needs to catch up with best practice. Again, a series of DEA models have been developed for these purposes since 1997. This route is now followed in most European countries. Effectively it means that real competition, which is not attractive since electricity distribution requires high infrastructure investments that should not be done in parallel by several firms in a given area, is substituted for by benchmarking; instead of competing directly, the DSOs compete via a benchmarking model.

1.2 Ideal evaluations

To see some of the difficulties and intricacies of benchmarking, let us start with a simple example. When we look at a firm or an organization, private or public, we are often interested to know how well it is doing. One way to think of this is illustrated in Fig. 1.1 below. We have a firm that has produced certain outputs using certain costs as indicated by the bullet in the output–cost diagram. The question is if this is a good performance?

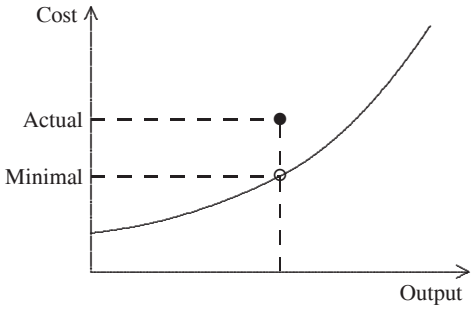


Fig. 1.1 An ideal evaluation

To evaluate the performance, we could use a cost function . It shows by definition the smallest possible costs of providing different output levels. If the cost function is as illustrated in Fig. 1.1, we can safely say that the firm has been inefficient. It is possible to produce the same outputs with less cost, or more outputs with the same cost, or some combination.

In fact, the excessive cost of the firm, i.e. the vertical distance between the actual cost level of the firm and the minimal necessary costs is an absolute measure of the inefficiency. The relative inefficiency could therefore be measure by

$$\text{InEfficiency} = \frac{\text{Actual cost} - \text{Minimal cost}}{\text{Actual cost}}.$$

The smaller the inefficiency, the better the performance.

Likewise, we could measure the relative efficiency directly as the ratio of minimal cost to actual costs

$$\text{Efficiency} = \frac{\text{Minimal cost}}{\text{Actual cost}} = 1 - \text{InEfficiency}.$$

The higher the efficiency, the better the performance.

We see that if we know the actual behavior of the firm, here represented by it is output and cost numbers, and if we have an appropriate model of the ideal performances, here represented by a cost function, we can easily make a performance evaluation. We could call this the *rational ideal evaluation* . It is a rational evaluation in the sense that we specify the preferences (e.g. to reduce costs) and possibilities (as given by the cost function), and we seek the best ways to pursue the goals. It is an ideal evaluation in the sense that we have all the relevant information.

More generally, the rational ideal evaluations can be described in the following way: From a standard microeconomic perspective, the performance of a firm is reflected in its ability to choose the best means (alternatives) to pursue its aims (preferences). In Fig. 1.2 we provide an illustration. The alternatives available is given by the technology T, here illustrated by the curved output isoquant. By definition the output isoquant shows the largest possible outputs for given inputs. The preferences given by a utility function $U(\cdot)$ is represented here by linear indifference curves. The indifference curves shows the outputs combinations that are equally good.

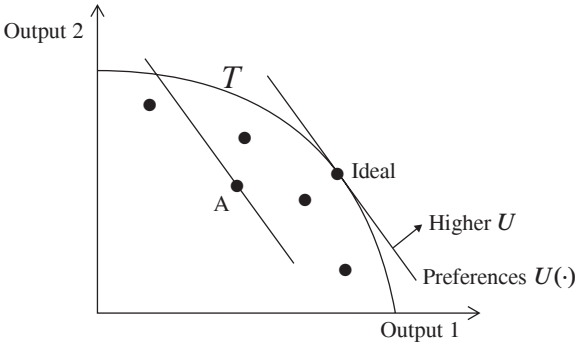


Fig. 1.2 Rational ideal set-up

The rational ideal performance evaluation would therefore compare the actually attained utility level with the maximally attainable utility level. For firm A this would be to compare $U(A)$ with $U(\text{Ideal})$. This would capture the effectiveness of firm A

$$\text{Effectiveness} = \frac{\text{Actual performance}}{\text{Ideal performance}} = \frac{U(A)}{\max_{y \in T} U(y)} = \frac{U(A)}{U(\text{Ideal})}$$

Note that we talk about effectiveness when we have an objective function and therefore can talk explicitly about goal attainment. When this is not the case and we rely on some proxy objective, we talk instead about efficiency.

In *real evaluations*, it is not entirely easy to apply this microeconomic cookbook recipe. In the *typical evaluation* we lack clear priorities U as well as clear information about the production possibilities T. In real evaluations, therefore, none of the elements in the rational ideal evaluation are known up front. Despite of this it is a useful conceptual idea. Basically, benchmarking is an attempt to approximate the economic idea of the rational ideal evaluation. We need to collect data to describe actual behavior, we need to estimate an approximation of the ideal relationship between inputs and outputs and we need to combine the actual performance with the ideal performance to evaluate the efficiency. Performance evaluation and efficiency analyses are basically concerned with these activities. *Benchmarking is a way to overcome these fundamental practical problems by moving from effectiveness to relative efficiency* . In the next sections, we explain the main steps used to accomplish this.

1.3 Key Performance Indicators and Ratios

A traditional way to overcome some of the difficulties of making rational ideal evaluations is to use what practitioners like to call *Key Performance Indicators, KPIs*. These are numbers that are supposed to reflect in some essential way the purpose of the firm.

Key Performance Indicators, KPIs, are widely used by firms, shareholders, regulatory agencies, researcher and others with an interest in performance evaluation. Most industries have very specific indicators reflecting the details of their technology. Network companies may for example consider maintenance costs per km of lines. The accounting literature has a series of financial indicators that are used across many industries to compare financial accounts. They include measures like Return on Assets (= net income/total assets), Gross Margin (gross profit/net sales), Debt Ratio (total liabilities/total assets), Price/Book (stock's capitalization/book values) to give a few examples.

As the examples illustrate, a Key Performance Indicator, KPI, if often a ratio of an output to an input. To see how this works, let us first assume that our firm only uses one input to produce one output. If we have input–output data from several firms, we can use this to determine who is doing best. We can simply compare what is often loosely called productivity , i.e. output per input. When we put the input–output combinations for each firm in a simple graph, it might look like the one in Fig. 1.3. Note that in this picture, we have inputs on the x–axis while we had inputs on the y–axis in the cost function example on page 6.

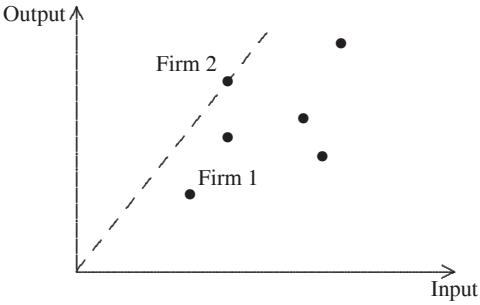


Fig. 1.3 Input–output combinations for several firms

The firm with the highest productivity is the one with the highest ratio of output per input, and that firm's level of productivity is shown as the slope of the dashed line in the figure. We can now compare the other firms to this special firm by comparing them to the dashed line.

We can call the ratio of productivity for firm 1 compared to firm 2 the efficiency ratio of firm 1 relative to firm 2, and we denote this ratio by E. Let (x^1, y^1) and (x^2, y^2) be the input–output combinations of the 2 firms. Then the productivities of the firms are $\frac{y^1}{x^1}$ and $\frac{y^2}{x^2}$, and the efficiency ratio E is

$$E = \frac{y^1/x^1}{y^2/x^2} = \frac{y^1/y^2}{x^1/x^2}.$$

It is important to observe that the traditional use of Key Performance Indicators, KPIs, are based on some *implicit assumptions*.

First, when we compare a firm with small output to a firm with large output in this manner, we implicitly assume that we can scale input and output in a linear fashion, i.e. we assume *constant returns to scale* . Thus even in this simple introductory example, our comparison depends on some assumptions. If we assume instead diminishing or increasing returns to scale our comparison might end up differently. In the ideal evaluation in Fig. 1.1 we did not make any such assumption—but then again, we presumed there that we knew the true relationship between input and outputs.

A second limitation of the KPI approach is that it typically involves only *partial evaluations* . One KPI may not fully reflect the purpose of the firm. We could have multiple inputs and therefore form several output–input ratios like above. We may for example be interested in the output per labor unit and the output per capital unit used in the production. Hence we would have two KPIs like in Fig. 1.4. The problem is now that the KPIs may not identify the same firm as the most productive firm. Firm 2 in Fig. 1.4 is having high output per labor unit but low output per capital unit, and firm 3 is having high capital productivity but low labor productivity. Of course, we could then say that firm 1 should strive to have the labor productivity of firm 1 and the capital productivity of firm 3. In many cases, however, this ideal is not feasible

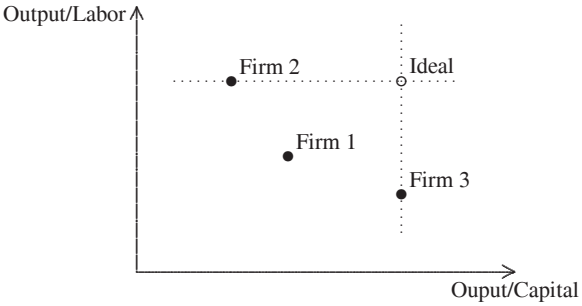


Fig. 1.4 Conflicting Key Performance Indicators (KPIs)

since there will be some substitution effect between labor and capital. Another way to put this is to say that partial benchmarks make misleading comparisons and that real firms may be compared to non-feasible, over-optimistic ideals.

A third and more intricate limitation of simple indicator approaches is known as the *Fox's Paradox* . It shows—in loose terms—that even if one firm are having higher values of all its partial productivity measures, it might have lower total productivity than another firm. The reason is that to do well in total, it is not only important to do well in the different sub-processes—it is also important to make use of the sub-processes that have relatively higher productivities than others. To illustrate this, consider the two firms in Table 1.1. Here we consider two firms serving patients

Table 1.1 Fox's paradox: Comparing Unit Cost

Firm	Cure UC	Care UC	Total UC
1	$\frac{10}{20} = .50$	$\frac{10}{40} = .25$	$\frac{20}{60} = 0.33$
2	$\frac{2}{3} = .66$	$\frac{21}{80} = .26$	$\frac{23}{83} = 0.29$

using either cure or (preventive) care. Firm 1 has provided cure to 20 patients at the costs of 10. Its cure unit cost, Cure UC, is therefore 10/20=.50. Similarly, firm 1 has provided care to 40 patients at the cost of 10. In total, they have therefore served 60 patients at a cost of 20. The interpretations of the numbers for firm 2 are similar. We see that the unit costs, UC, i.e. cost per unit served, is smaller in firm 1 for both cure and care. Still the total unit costs are higher than in firm 2. The reason is that firm 2 relies more on the relatively less costly treatment, care rather than cure.

We will now show how to overcome these limitations of KPIs. First we discuss how to relax the assumption of constant return to scale and next how to handle cases with multiple competing KPIs.

1.4 Technology and efficiency

When we compared input–output combinations for several firms in Fig. 1.3, we implicitly assumed that we can arbitrarily scale input and output up and down, i.e. we assumed constant returns to scale. We will now relax this assumption. We will look for an alternative definition of efficiency E with the same interpretation, but without this scaling assumption.

We will define *input efficiency* for an input–output combination (x, y) as the smallest factor E by which we can multiply the input x so that Ex can still produce the output y. If we were to use a smaller amount of input than Ex, it would be impossible to produce y. Hence

$$E(x, y) = \min\{\, e \mid ex \text{ can produce } y \,\}.$$

Another way to look at E is to say that it is possible to actually save $(1 - E)x$ of the input and still produce the same output y.

To determine whether or not an input can produce an output we need knowledge of the technology. For this purpose, we introduce the technology set.

The *technology set* T is the set of combinations of input and output such that the input can actually produce the output.

$$T = \{\, (x, y) \mid x \text{ can produce } y \,\}.$$

A main issue in benchmarking is to estimate what the technology set look like starting with some actual input–output observations from several firms.

If there is no noise in the data, then the technology set consists at the very least of our observations of input–output combinations for the observed firms; what they have produced it is apparently feasible to produce. The smallest technology set that contains all the observations is precisely the set of all the observations. But this is not an interesting technology set for further analysis as every new observation might bring another point or element into the set and thereby change it. We have therefore not done much in terms of modeling the technology. We want a technology set derived from the observations in such a way that not every new observation would lead us to expand it. To fulfill this, we will make assumptions about the technology set.

Our first assumptions is that we can dispense with any extra inputs and outputs: if an input–output combination is a feasible production for a firm, then any input–output where the input is larger and the output smaller is also a feasible production, i.e. is also in the technology set. We call this assumption *free disposability* . Starting from at set of observations numbered 1 to 6, the resulting free disposability technology set is the dotted area shown in Fig. 1.5.

A second common assumption is that if two input–output combinations are feasible productions, then any mixture of the two is also a feasible production. A mixture of two input–output combinations is called a convex combination, and we therefore talk about this as the *convexity* assumption, or we say that the technology set is

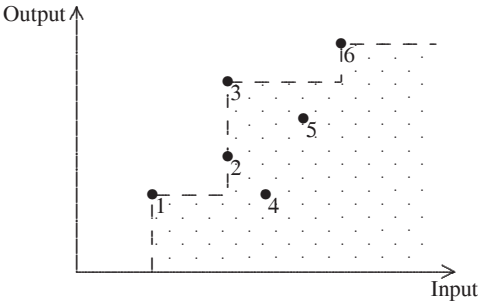


Fig. 1.5 Input–output combinations and free disposability

convex. If we assume both free disposability and convexity and have the six observations from Fig. 1.5, the technology set looks like in Fig. 1.6.

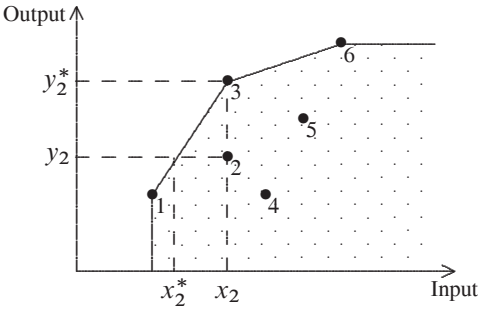


Fig. 1.6 Input–output combinations, free disposability and convex

When we looked at Key Performance Indicators (KPIs), we compared a firm with the best firm. With an estimated technology, we compare a firm with what is feasible given the technology set, i.e. we will compare it to the boundary or frontier of the technology set.

It can be seen from Fig. 1.6 that the frontier of the technology set is determined by the firms 1, 3, and 6, and therefore, we compare any input–output combination in the technology set with a mixture or convex combination of these firms.

Consider for example firm 2. We see that its input–output combination is an interior point in the technology set, and it is possible to produce y_2 by just using the input x_2^* instead of the observed input x_2 for unit 2. The *input efficiency* of firm 2 is therefore

$$E_2 = \frac{x_2^*}{x_2}$$

and we have $x_2^* = Ex_2$.

When we compare firm 2 with the boundary we really compare it with a convex combination of firms 1 and 3. Note also that the efficiency measured here is similar

to the efficiency calculated in relation to Fig. 1.1. The only difference is that we now use an estimated technology rather than one that was given a priori via a cost function.

If, instead, we keep the input for firm 2 fixed at x_2 then we can calculate the *output efficiency* by

$$F = \frac{y_2^*}{y_2}$$

such that $y_2^* = Fy_2$. For firm 2, using the input x_2 it would be technically possible to produce the output y_2^* and not just the smaller output y_2; i.e. the possible increase in output is $(F - 1)y_2$.

1.5 Many inputs and outputs

Let us now turn to the problem of multiple inputs and multiple outputs.

Most firms use multiple inputs to produce multiple outputs. If we can combine the inputs into one meaningful input aggregate, say costs, and all outputs into one meaningful output, say revenue, then we can safely use a single indicator as in Sect. 1.3 or—if we want to avoid the assumption of constant return to scale—a simple technology as in Sect. 1.4.

Unfortunately, such *aggregations* are often not possible. Just think a hospital using among others doctors and nurses to produce knee and heart operations. The aggregation of different treatments is certainly controversial, but also the aggregation of inputs may be problematic, e.g. if the labor market is not well-functioning.

Also, economist often refrain from measuring a systems (social) utility since it involves delicate problems of weighing together the utilities of individuals. In such cases, we are typically left with a multiple dimensional description of a systems end-user utilities, and further measurement problems arise. One system may be superior from the point of view of some individuals while another may be superior from the point of view of other individuals.

In the benchmarking literature, therefore, the idea of comparing single inputs to single outputs is abandoned. It is recognized that the multiple inputs and outputs exist and may interact and substitute for each others. Therefore, we use a more *system orientated* approach to the firm. A firm is seen as a transformation of resources into products and services. The transformation is affected by non-controllable variables as well as non-observable skills and efforts in the organization. The idea is now to measure the inputs, the outputs and the non-controllable variables and hereby to get an idea of the non-measurable managerial characteristics, the skills and effort as illustrated in Fig. 1.7 below.

In the evaluations, we shall therefore try to account for all the inputs, all the outputs and all the exogenous factors simultaneously. Only this way can we avoid the limitations of making partial evaluations.

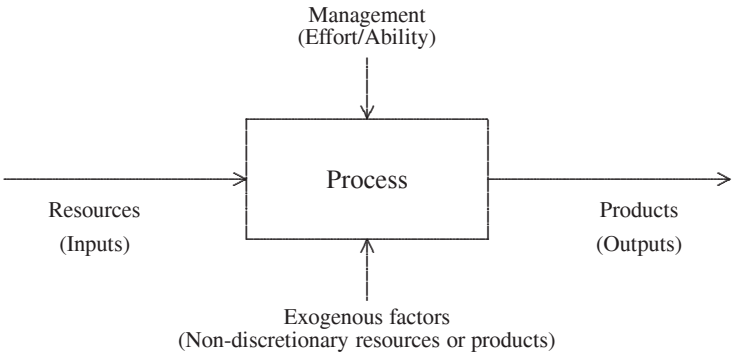


Fig. 1.7 Systems view

The systemic view, however, makes comparisons more complicated since we have to handle the multiple dimensions, and firms may be good in some dimensions and bad in others.

Let us consider some examples. In the case of two inputs we can draw the input isoquant for given outputs, and with two outputs we can draw the output isoquant or output frontier for given inputs as in Fig. 1.8. In the figure we have also marked an

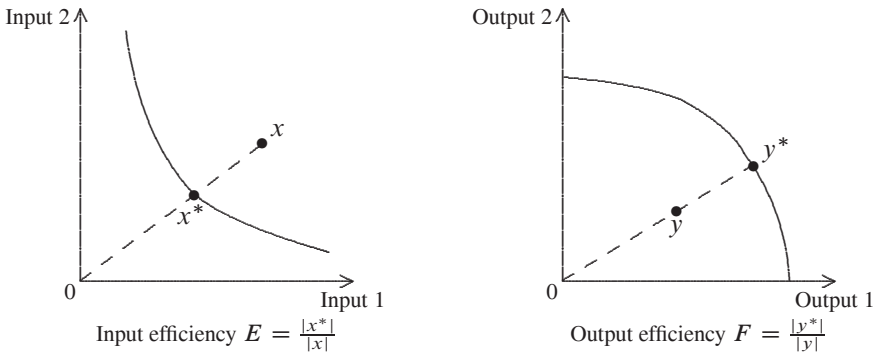


Fig. 1.8 Measuring Farrell efficiency

observed input combination x and an observed output combination y for which we want to calculate the efficiency.

It is clear in both cases that the firm is inefficient. It is possible to save inputs and still produce the same outputs since there are points to the south–west of x that are still above the input isoquant. Likewise, it is possible to expand the products and services represented by y since there are point to the north–east of y that are still

below the output isoquant (transformation curve). In other words, there are many possibilities to improve, and the question is how to summarize these possibilities.

To measure efficiency in such settings, the modern benchmarking literature has relied in particular on the Farrell (1957) measures . The idea of the Farrell measures is to focus on proportional changes—the same percentage reducing in all inputs or the same percentage increase in all outputs. Such changes corresponds to movements along the dashed lines in Fig. 1.8.

The Farrell input efficiency measures how much we can proportionally reduce the input and still produce the same output. The input efficiency is therefore calculated as the smallest number E such that $x^* = Ex$ where x^* is the point of intersection of the dashed line and the isoquant in the left part of Fig. 1.8. More formally, we have

$$\text{Farrell input efficiency} = E = \min\{\, e \mid ex \text{ can produce } y \,\} = \frac{|x^*|}{|x|}.$$

where $|x^*|$ is the length of the x^* vector, i.e. of the line between 0 and x^*. In the same way we can define the output efficiency as the largest factor that we can multiply on the output and still have a possible production for given input. The output efficiency is therefore calculated as the largest number F such that $y^* = Fy$ where y^* is the point of intersection of the dashed line and the transformation curve in the right part of Fig. 1.8. More formally, we have

$$\text{Farrell output efficiency} = F = \max\{\, f \mid x \text{ can produce } fy \,\} = \frac{|y^*|}{|y|}.$$

For inputs on or above the isoquant and ouptuts on or below the output isoquant curve we have $E \leq 1$ and $F \geq 1$, and the smaller is E and the larger is F, the more inefficient is the firm.

1.6 From effectiveness to efficiency

To summarize our discussions so far, modern benchmarking tools addresses two fundamental problems in practical evaluations. *The lack of clear preference or priority information is handled by moving from effectiveness to efficiency and the lack of a priori technological information is handled by making weak or flexible a priori assumptions, by estimating the technological frontiers, and by evaluating efficiency relative to the estimated frontier (best practices).*

An illustration of this is provided in Fig. 1.9 below. Here we consider hospitals that use the same resources to produce different combinations of operations. The data available in a given context is therefore represented by the dots. Now how can we evaluate a given hospital, say Hospital A.

We would ideally like to evaluate effectiveness as earlier, i.e.

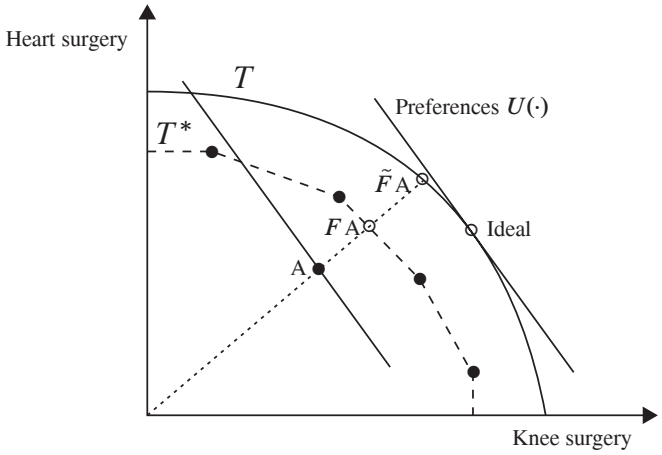


Fig. 1.9 A hospital example

$$\text{Effectiveness} = \frac{U(\text{A})}{\max_{y \in T} U(y)} = \frac{U(\text{A})}{U(\text{Ideal})}$$

However, this is not possible since we know neither U nor T.

The missing information about U is handled by focusing on efficiency instead of effectiveness. Hence, we ask how much we could have proportionally expanded A within the technology set T

$$\text{Absolute Farrell output efficiency} = \tilde{F} = \max\{ f \mid f\text{A} \in T \}$$

We see that A is projected into the point $\tilde{F}$A.

In reality, we also do not know T, so we approximate T by some set T^* determined from the performances of the other hospitals. In Fig. 1.9 we have used the traditional assumptions of free disposability and convexity. We now measure efficiency using the estimated technology T^* as our benchmark. We call this the relative efficiency since it is efficiency relative to the other hospitals in the sample as approximated by T^*, not relative to some absolute norm T, i.e.

$$\text{Relative Farrell output efficiency} = F = \max\{ f \mid f\text{A} \in T^* \}.$$

We see that A is now projected into the point FA.

We see also that $F \geq \tilde{F}$, i.e. relative efficiency is higher than absolute efficiency. This reflects that the relative performance standards are easier to live up to than the absolute ones. We can also say that we handle the lack of information about the underlying true technology T by taking *a cautious approach to the estimation of the technology which in turn lead to a cautious (higher) estimate of efficiency*.

A large part of this book is focusing on these two steps. We shall look for different ways to measure efficiency as opposed to effectiveness, and we shall look for

different ways to estimate a technology from observed performances. We cover in more depth the basics of technology estimation in Chap. 3 and the basics of efficiency measurement in Chap. 2. We then turn to more advanced frontier estimation methods and applications in the subsequent chapters. A first survey of the more advanced frontier estimation approaches is provided in the next section.

1.7 Frontier models

Modern benchmarking analyses increasingly use *best practice* or frontier analysis methods. The idea is to model the frontier of the technology rather than to model the average use of the technological possibilities. This has methodological as well as practical advantages. One very practical aspect is that it is often more interesting to learn from the best than to imitate mediocre performances.

Frontier analyses in general and Data Envelopment Analysis (DEA) and Stochastic Frontier Analysis (SFA) methods in particular, are developing rapidly in theory as well as in practice, and they will be our main focus of attention in the remainder of this book. Before digging into the details of these methods, however, we give a non-technical survey of the methods and their main pros and cons in this section.

1.7.1 A simple taxonomy

In the benchmarking literature—as in traditional statistical literature—it is common to distinguish parametric and nonparametric approaches. *Parametric models* are characterized by being defined a priori except for a finite set of unknown parameters that are estimated from data. The parameters may for example refer to the relative importance of different cost drivers or to the parameters in the possibly random noise and efficiency distributions. *Nonparametric models* are characterized by being much less restricted a priori. Only a broad class of functions, say all increasing convex functions, or even production sets with broadly defined properties, are fixed a priori and data is used to estimate one of these. The classes are so broad as to prohibit a parameterization in terms of a limited number of parameters, and they can therefore be termed non-parametric.

Another relevant distinction is between deterministic and stochastic models. In *stochastic models* , one make a priori allowance for the fact that the individual observations may be somewhat affected by random noise, and tries to identify the underlying mean structure stripped from the impact of the random elements. In *deterministic models* , the possible noise is suppressed and any variation in data is considered to contain significant information about the efficiency of the firms and the shape of the technology.

The two dimensions leads to a 2×2 taxonomy of methods as illustrated in table 1.2 on the following page. A few seminal references are included. We emphasize

Table 1.2 A taxonomy of frontier methods

	Deterministic	Stochastic
Parametric	*Corrected Ordinary Least Squares (COLS)*	*Stochastic Frontier Analysis (SFA)*
	Aigner and Chu (1968), Lovell (1993), Greene (1990, 2008)	Aigner et al (1977), Battese and Coelli (1992), Coelli et al (1998a)
Non-parametric	*Data Envelopment Analysis (DEA)*	*Stochastic Data Envelopment Analysis (SDEA)*
	Charnes et al (1978), Deprins et al (1984)	Land et al (1993), Olesen and Petersen (1995), Fethi et al (2001)

that for each class of model, there exist a large set of model variants corresponding to different assumptions about the production technology, the distribution of the noise terms etc.

To illustrate the differences, consider a simple cost modeling context. In this setting, we seek to model the costs that results when best practice is used to produce one or more outputs. We have data from a set of production units as indicated in Fig. 1.10.

Now, *Corrected Ordinary Least Squares (COLS)* corresponds to estimating an ordinary regression model and than making a parallel shift to make all firms be above the minimal cost line. *SFA* on the other hand recognizes that some of the variation will be noise and only shift the line—in case of a linear mean structure—part of the way towards the COLS line. *DEA* estimates the technology using what is known as the minimal extrapolation principle . It finds the smallest production set (in the illustration the set above the DEA curve) containing data and satisfying a minimum of production economic regularities. Assuming free disposability and convexity, we get the DEA model illustrated in Fig. 1.10. Like COLS, the DEA cost function it is located below all cost–output points, but the functional form is more flexible and the model therefore adapts closer to the data. Finally, *Stochastic Data Envelopment Analysis (SDEA)* combines the flexible structure with the possibility, that some of the variations in data may be noise, and only requires most of the points to be enveloped.

In Fig. 1.10 we have included a fifth frontier, termed *Engineering*. The idea is to base the modeling on data from engineers about best possible performance, perhaps in idealized settings. We will discuss engineering approaches in some examples in this book, but since the approaches differs with the application area, no general outline will be given.

1.7.2 Pros and cons

In the choice between DEA and SFA, a key question is whether one wants flexibility in the mean structure or precision in the noise separation.

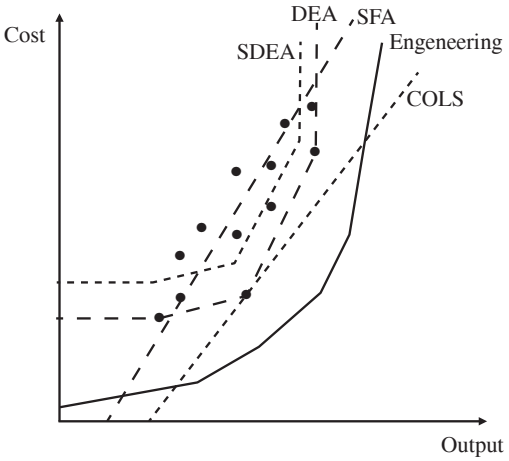


Fig. 1.10 Alternative frontiers

An important property of a benchmarking approach is its ability to reflect and respect the characteristics of the industry. This requires that it is a *flexible model* in the sense that its shape (or its mean structure to use statistical terms) is able to adapt to data instead of relying excessively on arbitrary text book assumptions. This is particular important in attempts to support learning, individual motivation and coordination. It is probably less important in models aimed at evaluating system wide shifts, e.g. the impact of some policy change. The nonparametric models are by nature superior in terms of flexibility.

Another important property of a benchmarking approach is its ability to cope with *noisy data*. A robust estimation method gives results that are not too sensitive to random variations in data. The stochastic models are particularly useful in this respect.

In summary, the nonparametric models are the most flexible in terms of the production economic properties that can be invoked while the stochastic models of course are the most flexible in terms of the assumptions one can make about data quality.

Ideally, then, we would like to use flexible models that are robust to random noise. This would favor SDEA models. The problem however is the properties come at a cost. The estimation task become bigger, the data need larger and still we cannot avoid a series of strong assumptions about the distributions of the noise terms. Coping with uncertainty requires us to dispense somewhat with flexibility and vice versa. We would furthermore argue that a lack of stochasticity can be partly compensated by a flexible mean structure—and a restricted mean structure can be somewhat compensated by allowing for random elements. This means that DEA and SFA are very useful methods and that we do not necessarily need to move to SDEA.

Besides these pros and cons, it is worthwhile to mention a general advantage of DEA and SFA models compared to earlier and less advanced benchmarking meth-

ods. Both methods require no or very little preference, price or priority information and are able to cope effectively with multiple inputs and outputs. This is most likely the general reason they have become so popular and may partially explain that several thousand scientific papers and reports have used these methods since they were first introduce some 30 years ago.

1.8 Software

There are by now several free software packages as well as commercial programs that are useful when the methods of this book are put to work on real data sets. We shall not cover these in any details, but we shall illustrate several of our models and methods by showing how they can be implemented in one particular general purpose, very powerful and free software called R. We provide a brief introduction to this software in Appendix A.

In parallel to writing this book, we have developed our own R routines, an R-package called *Benchmarking* , that makes it easy to implement both the standard models that other software can handle as well, and other variants that users may find interesting. Our aim with this R package is to make applications very easy without limiting the variations of models and calculations that innovative researchers and practitioners would seek to implement to get benchmarks that fit their particular purpose the best.

1.9 Summary

The aim of this chapter has been to give an elementary and intuitive introduction to the subject of this book.

Benchmarking was defined as relative performance evaluation of firms (or other production entities) that transforms the same types of inputs (resources) into the same type of outputs (services). This is useful in many contexts and can facilitate both learning, coordination and motivation.

Benchmarking is simple if one can aggregate the objective of the firm into a single criterion, a utility function, and if one has a perfect description of the possibilities, the technology. In such contexts one can make a rational ideal evaluation by comparing the attained utility value with the maximal value that is possible to obtain in the technology set. Unfortunately, in real evaluations, one typically lacks information about both the overall objective of the firm and of its possibilities. Benchmarking is a way to overcome these fundamental practical problems by moving from effectiveness to relative efficiency.

A common approach in practice in to define one or more Key Performance Indicators (KPIs) and to compare these among the firms. While this is useful in very simple cases, this approach has the drawbacks that it presumes constant return to

scale and that different KPIs may point to different ideal firms and that the combination of these typically is not feasible but rather too optimistic.

The approaches to benchmarking that we cover in this book therefore take a different approach/view. We start from a system description of all inputs used, the outputs produced and the contextual characteristics possibly affecting the transformation of input to output. We estimate the underlying technology using systematic assumptions and methods and we measure how much a given firm can improve by moving to the frontier of the estimated technology. The technologies we estimate are usually presumed to have certain mild regularities like free disposability and convexity, but otherwise the aim is to let the data define the technology to the largest possible extent. The improvement possibilities can be captures in many different ways but the most common approach to get a single efficiency measure is to rely on the Farrell idea of proportional improvements in all inputs (or all outputs).

In a survey of methods, one can distinguish between parametric and non-parametric methods and between stochastic and non-stochastic methods. The two approaches that we shall mainly cover in this book are the non-parametric, deterministic approach called Data Envelopment Analysis (DEA) and the parametric, stochastic approach called Stochastic Frontier Analysis (SFA). They both enable us to work with multiple inputs and outputs, and hereby to perform comprehensive evaluation of many different production entities, including non-for profit firms and public organizations. DEA is advantageous by having a very flexible production structure while SFA is advantageous by allowing a better separation of noise and inefficiency.

1.10 Bibliographic notes

Since they were first introduced some 30 years ago, the DEA and SFA methods have become extremely popular and several thousand papers have been produced extending and applying these methods. For a textbook introduction to DEA, see Charnes et al (1995) or Cooper et al (2007). A popular introduction to both DEA and SFA is Coelli et al (1998a).

The learning perspective using interactive benchmarking was first introduced (under the name of internet based benchmarking) in Bogetoft and Nielsen (2005). Software to support such exercises has since been developed and used in several industries and countries, cf. e.g. www.ibensoft.com.

The coordination and reallocation perspective was early introduced into the DEA literature. Lewin and Morey (1981) for example discuss the decomposition of inefficiency in a hierarchical organization, and Brännlund et al (1995) and Brännlund et al (1998) study the Swedish pulp and paper industry using a DEA model. They estimate the cost of the existing transmission constraints at the individual firms and the gains from reallocation. Extensions of these ideas to evaluate the possible gains from reallocating fishery quota and agricultural production rights are given in Andersen and Bogetoft (2007) and Bogetoft et al (2007a), respectively. The sugar beet

exchange is described in Bogetoft et al (2009). An variant of these ideas is used to evaluate mergers in Chap. 9.

The motivation perspective has always been implicitly available in the benchmarking literature. The formal modeling of how to tie budgeting, salary, allowed revenues etc to the outcome of the benchmarking has however been more limited. Notable early contributions are Banker (1980) and Banker et al (1989) while later contributions include Dalen (1996); Dalen and Gomez-Lobo (1997, 2001); Resende (2001); Sheriff (2001); Thanassoulis (2000); Wunsch (1995). Explicit combinations with agency models has been initiated in a series of papers by Bogetoft and coauthors, cf. e.g. Bogetoft (1990, 1994b,a, 1995, 1997, 2000), Agrell and Bogetoft (2001a), Agrell et al (2002), Agrell et al (2005b). Some of these models, in particular Bogetoft (1997), has subsequently been implemented in European regulation systems as we shall discuss in Chap. 10.

The Fox Paradox and was first discussed in the literature by Fox (1999).

Chapter 2
Efficiency Measures

2.1 Introduction

In Chap. 1, we introduced efficiency as the use of the fewest inputs (resources) to produce the most outputs (services). This idea is fundamental to much of modern benchmarking literature because it allows us to evaluate performance without clearly defined preferences. That is, we avoid the difficult task of estimating preference functions and deciding on exact priorities. We will expand on this below.

Although the notion of efficiency is simple and intuitive at first glance, there are actually many different ways to conceptualize efficiency. We shall discuss some of the most common concepts in this chapter. We will cover classical concepts from production theory, including technical efficiency, allocative efficiency, and scale efficiency, as well as more advanced concepts like dynamic efficiency and structural efficiency.

Moreover, several of these concepts can be operationalized in different ways. We can, for example, measure technical efficiency in terms of input space, output space, or both types of spaces. We can also measure it in specific directions, etc.

The aim of this chapter is to provide an overview of efficiency-related concepts as well as bits and pieces of the relevant theoretical background.

2.2 Setting

In pursuing this aim, we will generally assume that the technology is given. We focus on a given firm and can therefore describe the setting in the following way: A firm k has used m inputs $x^k = (x_1^l, \ldots, x_m^k) \in \mathbb{R}_+^m$ to produce n outputs $y^k = (y_1^k, \ldots, y_n^k) \in \mathbb{R}_+^n$. The set of feasible production plans or input-output combinations available to firm k is given by the technology or production possibility set T,

$$T = \{\, (x, y) \in \mathbb{R}_+^n \times \mathbb{R}_+^m \mid x \text{ can produce } y \,\}.$$

P. Bogetoft and L. Otto, *Benchmarking with DEA, SFA, and R*, International Series in Operations Research & Management Science 157, DOI 10.1007/978-1-4419-7961-2_2,

There are many ways to construct the technology T. We have already illustrated some of these methods in Chap. 1, and we will take a closer look at the basic assumptions that one can make about technologies in Chap. 3. Moreover, we shall spend much of the book describing alternative methods like Data Envelopment Analysis DEA and Stochastic Frontier Analysis SFA, which involve the construction of technologies based on actual observations. For now, however, it does not matter how we estimate T. The same efficiency concepts are applicable to technologies estimated in different ways.

2.3 Efficient production

Efficiency is generally a question of using few inputs (resources) to produce many outputs (services).

To be more precise, let us consider two firms, (x^1, y^1) and (x^2, y^2). We say that firm 2 dominates or is more efficient than firm 1 if it uses no more inputs to produce no fewer outputs and is doing strictly better in at least one dimension.

Dominance. (x^2, y^2) dominates (x^1, y^1) if and only if $x^2 \leq x^1$, $y^2 \geq y^1$, and $(x^1, y^1) \neq (x^2, y^2)$

Note that we require the dominating firm, firm 2, to use no more inputs to produce no less outputs than firm 1 and to not be exactly similar to firm 1. Therefore, we require the dominating firm to be strictly better in at least one dimension (to use strictly less of an input or produce strictly more of an output).

Dominance allows us to partially rank firms. Some firms can be compared, while others cannot. This is illustrated in the left panel in Fig. 2.1. firm 2 dominates firm 1, while firm 3 neither dominates nor is dominated by firm 1 or firm 2.

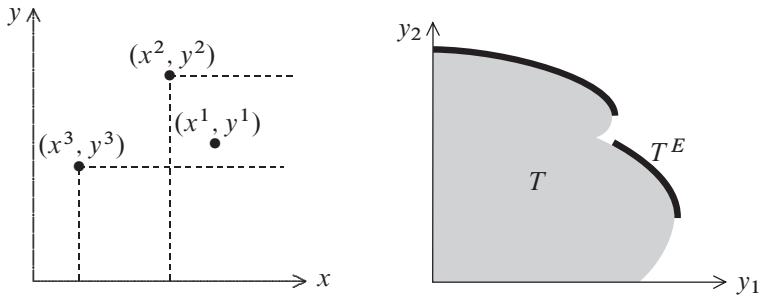


Fig. 2.1 Dominance and efficiency

The dominance relationship is relevant because almost everyone would prefer a more efficient or dominating production plan to the less efficient one that it dominates. For this to hold, we need our preference to be increasing in outputs and

decreasing inputs. Thus, dominance is a weak expression of preferences that allows us to partially rank production plans.

In economics, the efficient firms are those that cannot be dominated by other firms. To determine which firms are efficient, we thus need a description of all possible firms (e.g., a listing or a technology set). For a given technology set T, we define efficiency as follows:

Efficiency. (x, y) is efficient in T if and only if it cannot be dominated by some $(x', y') \in T$.

The efficient subset of T, T^E is

$$T^E = \{ (x, y) \in T \mid (x, y) \text{ is efficient in } T \}.$$

The efficient subset T^E of T are the inputs-output combinations that cannot be improved. They represent best practices. An illustration is provided in the right panel of Fig. 2.1. Here the technology set T for 2 outputs is the shaded area, and the efficiency set T^E is the bold part of the frontier. In the production economics literature, this notion of efficiency is sometimes called *Koopmans-efficiency* to distinguish it from other types of efficiency.

The focus on efficiency is natural from a *theoretical perspective*. On the one hand, efficiency is not too strong a requirement; under conditions of mild regularity, one can always identify an optimal production plan from among the efficient ones. On the other hand, we cannot generally strengthen the efficiency requirement; any efficient plan may be the uniquely optimal plan given perfectly sensible underlying but unknown preference functions.

The focus on efficiency is also convenient from an *applied perspective*. One of the main obstacles to the evaluation of effectiveness is to select the objectives or preferences against which we should gauge performance. Here, efficiency provides an easy way out because it only requires that more outputs and fewer inputs are preferable. Thus, instead of engaging in dead-end discussion about overall objectives, we create a partial ranking that will be agreed on by almost everyone. It is worth remembering, however, that this logic also means that while *efficiency is a necessary condition for effectiveness, it is not a sufficient one*. In fact, in terms of a particular technology, an inefficient firm may well be better than a fully efficient one. We could rephrase this by saying that it is not sufficient to run fast; it is also important to run in the correct direction—and it may be better to run at a moderate speed in the right direction than at full speed off-course.

So far, we have defined and explained the relevance of efficiency. We have focused on which firms are efficient and which are not. Additionally, we have introduced a partial ranking of firms in terms of dominance. In the following sections, we will study how to measure efficiency levels. We want to go beyond the efficient/inefficient dichotomy and measure degrees of (in)efficiency.

2.4 Farrell efficiency

The single most widely used approach to measuring the degree of efficiency in a general multi–input and multi–output setting is the strategy suggested by Debreu and Farrell, usually referred to simply as Farrell efficiency. The idea is to ask if it is possible to reduce the input without changing the output. Seeking to process multiple inputs and outputs in a simple way, we look for a proportional reduction of all inputs.

The *input–based Farrell efficiency* or just *input efficiency* of a plan (x, y) relative to a technology T is defined as

$$E = \min\{\, E > 0 \mid (Ex, y) \in T \,\}$$

i.e., it is the maximal proportional contraction of all inputs x that allows us to produce y. Thus, if $E = 0.8$, it indicates that we could have saved 20% off all inputs and still produced the same outputs.

Likewise, output–based Farrell efficiency or *output efficiency* is defined as

$$F = \max\{\, F > 0 \mid (x, Fy) \in T \,\}$$

i.e., the maximal proportional expansion of all outputs y that is feasible with the given inputs x. Thus, a score of $F = 1.3$ suggests that we could expand the output by 30% without spending additional resources.

A small-scale example of this concept using one input and one output is provided in Fig. 2.2. We see that we can reduce input x to x^* without losing output and that

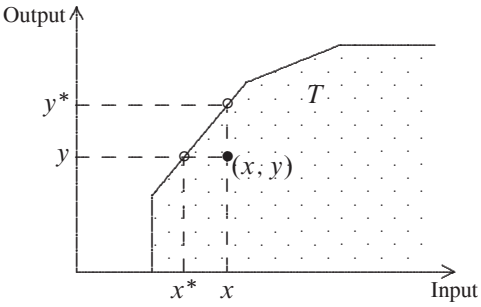


Fig. 2.2 Farrell efficiency in one–input/one–output example

we can increase output y to y^* without using more resources. Therefore, we have

$$E = \frac{x^*}{x},$$
$$F = \frac{y^*}{y}.$$

Table 2.1 Two–input, two–output example

Firm	Input A	Input B	Output C	Output D
1	10	20	20	20
2	20	10	40	20
3	20	30	60	80
4	30	30	80	60
Our	30	20	36	10

Figure 2.3 illustrates how Farrell efficiency is calculated when there are two inputs and two outputs. In the left panel, we show the input isoquant corresponding to the output level y that our firm is producing, and in the right panel, we show the output-isoquant corresponding to the inputs x that our firm is using.

Proportional reduction and expansion correspond to movements along the dashed lines in the two panels. Input efficiency is therefore calculated as the smallest number E that we can multiply on x and remain on or above the isoquant. Likewise, output efficiency is calculated as the largest number F that we can multiply on y and remain below or at the output isoquant. For inputs above and on the input isoquant and outputs below and on the output isoquant curve, we have $E \leq 1$ and $F \geq 1$. The smaller E is and the larger F is, the less efficient the firm is.

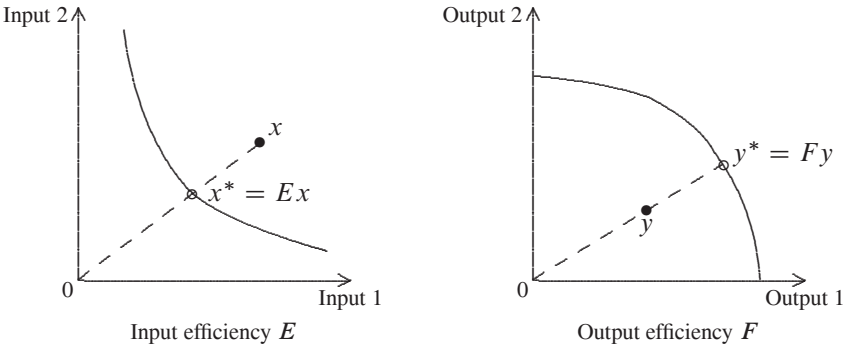


Fig. 2.3 Farrel efficiency

Numerical example

To better understand the logic of the Farrell measures, consider an example in which the technology T is formed using free disposability on the four firms in the upper part of Table 2.1. By this, we mean that any production plan dominated by one of our observed plans is feasible. We are interested in evaluating "our" firm as given in the last row.

Now we need to look for firms that are using fewer inputs to produce more outputs than our firm. In terms of input usage, we see that only firms 1 and 2 qualify because firms 3 and 4 use too much of input B. At the same time, we see that that firm 1 is not producing enough of output C but that firm 2 produces enough of both outputs. Thus, in effect, firm 2 is really the only firm we can compare with using only dominance (or free disposability).

Now consider the input efficiency of our firm compared to that of firm 2. We see that compared to firm 2, our firm could reduce input A with a factor 20/30 and input B with a factor 10/20. Because it has to decrease in both dimensions, the smallest feasible reduction factor is 20/30. Therefore,

$$E = \frac{20}{30} = 0.67.$$

In a similar way, we see that in terms of output C, our firm could expand with a factor 40/36 by imitating firm 2, and in terms of output D, it could expand with a factor 20/10. Again, because we are looking for the largest possible expansion that works in all output dimensions, we must settle on an expansion of 40/36, i.e.,

$$F = \frac{40}{36} = 1.11.$$

These results are illustrated in Fig. 2.4. In particular, we see that on the input side, it is input A that becomes binding, whereas, on the output side, it is output C that becomes the limiting product.

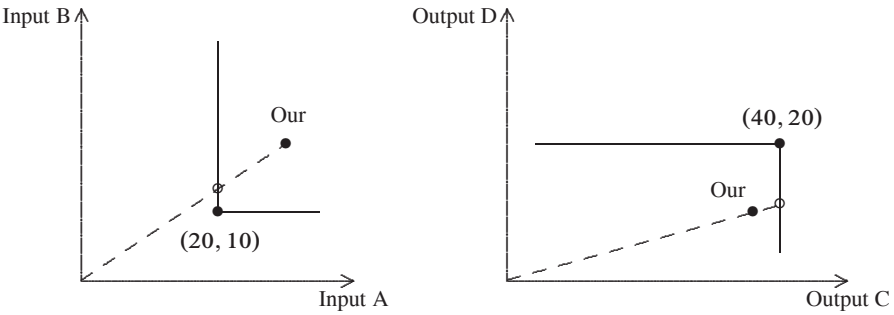


Fig. 2.4 Illustration of numerical example

2.4.1 Non-discretionary inputs and outputs

In applications, we often have situations in which some of the inputs or some of the outputs are fixed and uncontrollable, at least in the short run or using the discretionary power of the firm or unit that we seek to evaluate. A very simple but useful way to handle such situations is to only look for improvements in the discretionary (controllable) dimensions. In this way, if we divide the input and outputs into variable (v) and fixed (f) inputs and outputs as in $(x, y) = (x_v, x_f, y_v, y_f)$, we can define the input and output variants of the Farrell measures as follows:

$$E^* = \min\{\, E > 0 \mid (Ex_v, x_f, y_v, y_f) \in T \,\}$$
$$F^* = \max\{\, F > 0 \mid (x_v, x_f, Fy_v, y_f) \in T \,\}.$$

E^* indicates that we can proportionally reduce all variable inputs x_v with a factor E^* without using more of the fixed inputs x_f and without producing fewer of the outputs $y = (y_v, y_f)$. Likewise, the interpretation of F^* is that we can proportionally expand all variable outputs y_v without reducing any of the fixed outputs y_f and without using more inputs than $x = (x_v, x_f)$.

2.4.2 Using Farrell to rank firms

The main outcome of many benchmarking studies is a list of the Farrell efficiency values of the firms in an industry. Such lists or league tables are studied with interest and care because they are often considered to provide firm rankings with the best firms having the largest E and the worst having the lowest E (or vise versa for F).

One can discuss whether efficiency measures can really be used to rank firms or whether they solely provide individual measures of efficiency and thus improvement potential.

Purists would argue that rankings using Farrell efficiency are only possible to a very limited degree. A case can be made only for comparing firms where one dominates the other. In such situations, the efficiency score achieved by comparing one to the other is simply a way to quantify the amount of dominance.

One can also take a more pragmatic view and argue that even in cases in which the two units are not comparable based on dominance, the Farrell measure still provides a performance measure, and that low Farrell efficiency is an indication of high improvement potential. It is important to note, however, that this is not a simple ranking in which everyone is competing with everyone. It is more similar to a golf tournament or a horse race with handicapping. The technology defines the performance standard for each of the firms, and only hereby can we compare firms that produce different service mixes or use different input mixes.

Still, this use of efficiency scores presumes that the inputs and outputs correctly characterize the available options and that we do not have any more information about the relative importance of the different inputs and outputs.

Additionally, in reality, the technology may be described more precisely in some parts of the production space than others, and we shall talk at some length about bias in DEA in Chap. 5. Such biases in our descriptions make it less reasonable to consider the ranking as the result of a fair race. Indeed, in such cases, one can argue that not all firms participate in the same race and that the rankings are not fair because it is easier to be close to the winner in some races (with fewer or less talented competitors) than in other races. These differences make it difficult to make comparisons across races. (The second-best runner in Denmark is probably quite a bit worse than the second-best US runner).

It is also worthwhile to note that Farrell efficiency is not exactly the same as traditional (Koopmans) efficiency as introduced in Sect. 2.3. That is, $E(x, y; T) = F(x, y; T) = 1$ does not imply $(x, y) \in T^E$. This situation occurs when some inputs can be reduced and/or some outputs can be expanded individually but there is no option to contract or expand all inputs or outputs simultaneously (i.e., when we are on a horizontal or vertical part of the isoquants). This is one drawback of the Farrell measure.

2.4.3 Farrell and Shephard distance functions

Farrell efficiency depends on our starting point (x, y) and the technology set T. Instead of using E and F above, it would thus be more precise to use the longer notation $E((x, y); T)$ and $F((x, y); T)$. In many contexts, however, this would be too cumbersome and we simply use E and F or perhaps $E(x, y)$ and $F(x, y)$. In some cases, we also call these efficiency measures *distance functions* or, more precisely, input distance functions and output distance functions. This nomenclature emphasizes that they are not just numbers but are also procedures (functions) that map technologies and observations into real numbers.

Some prefer to work with the so-called Shephard measures rather than the Farrell measures. For the sake of completeness, we note that the *Shephard distance functions* are simply the inverse of the Farrell ones,

$$D_i(x, y) = \max\{\, D > 0 \mid \left(\frac{x}{D}, y\right) \in T \,\} = \frac{1}{E(x, y)}$$

$$D_o(x, y) = \min\{\, D > 0 \mid \left(x, \frac{y}{D}\right) \in T \,\} = \frac{1}{F(x, y)}.$$

The function D_i is called the (Shephard) input distance function and D_o the (Shephard) output distance function. Some computer programs calculate these functions rather than the Farrell variants.

2.5 Directional efficiency measures

In the Farrell (and Shephard) approach to efficiency measurement, all inputs are reduced or all outputs are expanded by the same factor. This proportional adjustment has been challenged by a series of alternative efficiency measurements approaches. We cover a few of these here.

An early suggestion in the DEA literature was to consider simultaneous improvements on the input and output side by basically combining the Farrell input and output efficiency measures into one measure, sometimes referred to as the *graph hyperbolic measure of technical efficiency*

$$G = \min\{\, G > 0 \mid (Gx, \frac{1}{G}y) \in T \,\}.$$

In G, we seek to simultaneously reduce inputs and expand outputs as in the Farrell approach. The input side is exactly as in the E measure, and the output side is in the spirit of the F measure; when we reduce G, we expand $1/G$, which is like the F factor is in the Farrell output efficiency measures. Also note that for $(x, y) \in T$, we have $G \leq 1$.

The interpretation of a graph hyperbolic efficiency G is that we can make due with input Gx and simultaneously expand output to $\frac{1}{G}y$. This is illustrated in Fig. 2.5. The curve traversed by $(Gx, \frac{1}{G}y)$ when G takes all positive values is a hyperbola; this is indicated by the dashed line, and by comparing the intersection and the original point, we can measure G on either the input or the output axis as indicated in the figure.

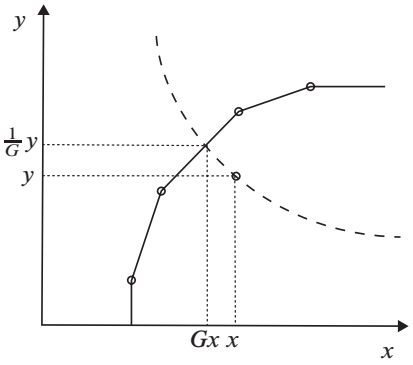


Fig. 2.5 Graph efficiency measure

In applications, G is not always easy to implement because of the non-linearities involved.

A more profound alternative or generalization of Farrell's proportional approach is based on *directional distance functions*. We will discuss this approach now, and to simplify the exposition, we initially focus on the input side.

The purpose of directional distance functions is to determine improvements in a given direction $d \in \mathbb{R}^m_+$ and to measure the distance to the frontier in such d-units. This process leads to a directional distance or *excess function*

$$e = e(x, y; T, d) := \max\{ e \in \mathbb{R}_+ \mid (x - ed, y) \in T \}.$$

The excess $e(x, y; T, d)$ has a straightforward interpretation as the number of times the input bundle d has been used in x in excess of what is necessary to produce y. Therefore, a large excess reflects a large (absolute) slack and a considerable amount of inefficiency. It shows how many times we can harvest the improvement bundle d if we were to learn best practice.

An illustration is provided in Fig. 2.6 for 2 different directions $(1, 0.25)$ and $(.25, 4)$ in addition to the usual Farrell direction. In the figure, we have also indicated the projection points using circles. On this basis, the efficiency figures can be calculated.

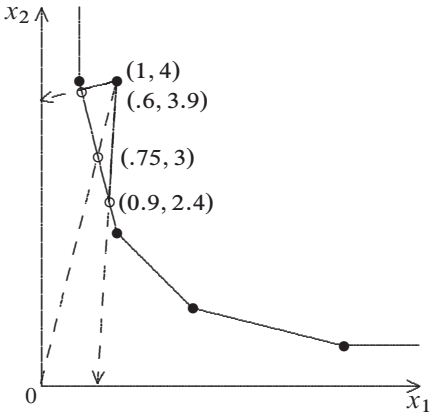


Fig. 2.6 Input directional efficiency

The directional distances and the Farrell input efficiency in this case become

$$\begin{aligned} e((1,4);(1,0.25)) &= 0.48 \\ e((1,4);(0.25,4)) &= 0.39 \\ E((1,4)) &= 0.75 \end{aligned}$$

We note that the directional distances are not comparable across different directions. Excess values depend on the directions in which we move, as this example shows. This dependence means that we need to exercise care in interpreting the results and particularly in comparing excess values across firms and directions. On the other hand, it is also a useful property because by measuring the excess in different directions, one can get a picture of which particular resources a firm seems to have

in excess. This approach is sometimes called multidirectional efficiency analysis, MEA, and can be used to select improvement strategies, etc.

The excess values also depend on the length of the direction vector. Thus, for example, if we double the length of the improvement direction d, the number of times we can save the doubled vector is halved. More generally, for arbitrary $\theta \geq 0$, we have

$$e((x, y); \theta d) = \frac{1}{\theta} e((x, y); d).$$

Again, this simply requires us to be explicit in interpreting the results and making comparisons across different firms and in different directions.

The Farrell approach is, in principle, just a special variant of the directional distance function approach, where we use the firms own inputs as the direction vector. Thus, it is straightforward to see that

$$e((x, y); x) = 1 - E(x, y)$$

That is, with direction equal to what is present in the existing input production plan, the excess function is a measure of the inefficiency of the firm as determined using the Farrell method. If Farrell efficiency is 80%, for example, the excess is 20%. Likewise, the Farrell efficiency measures when some of the inputs or outputs are fixed are special variants of the directional distance approach. We have, for example,

$$e((x, y); (x_v, 0)) = 1 - E^*(x_v, x_f, y).$$

Rather than creating a direct dichotomy between controllable and non-controllable elements, the directional distance function approach allows us to work with grades of discretion—some dimensions can be controlled more easily than others, and some dimensions are more desirable to change than others.

Like in the graph efficiency, we can combine the input and output efficiency perspectives using the direction distance function approach. That is, we can examine whether it is possible to use fewer inputs and produce more outputs. Thus, we can look for changes in the direction $(d_x, d_y) \in \mathbb{R}^m_+ \times \mathbb{R}^n_+$ and define the directional excess e as

$$e = \max\{ e > 0 \mid (x - ed_x, y + ed_y) \in T \}.$$

With one input and one output and the direction $(d_x, d_y) = (1, 1)$, we have the situation indicated in Fig. 2.7, where the arrow indicates the direction.

An important question, in theory as well as in practice, is which direction is best. The correct (but also somewhat easy) answer at this stage is that it depends on the application. We will return to this question in Sect. 2.9 below because it is also related to the question of using input– or output–based efficiency measures, technical or allocative efficiency measures, etc.

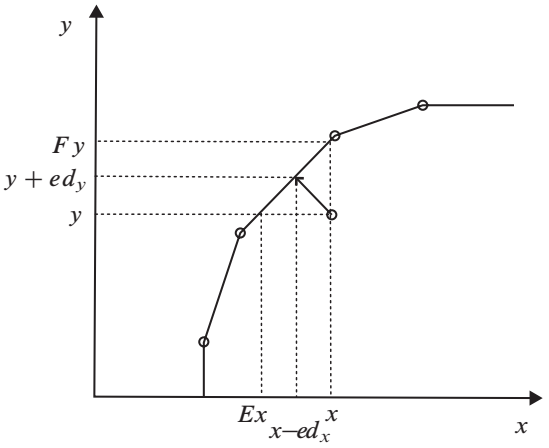


Fig. 2.7 Input and output directional efficiency

Practical application: Benchmarking in waterworks

Rather than theoretically discussing the pros and cons of different directions, one can take a pragmatic approach and see the direction as a steering instrument by which a user can control the projection of a firm on the efficient frontier. This approach in used in the interactive benchmarking system called IBEN that is used by Danish Waterworks, cf. also Sect.1.1.1, to support individual learning among waterworks. An illustration is provided in Fig. 2.8. In the specific model (technology), it is presumed that the evaluated waterworks have used two inputs to produce two outputs. The inputs are the Opex (DC1000) and Capex (DB1750) measures, and the outputs are the water distributed (DA1300) and length of the water lines (DA0320). We see, for example, that Hørsholm has used 3.43 million DKK of Opex and 2.3 mio DKK of Capex to distribute 1.362 million m^3 of water and maintain 149 km of waterlines. 1 million DKK is approximately 150 thousand Euro.

In the illustration, we see that the user has chosen to look for improvements in all directions (i.e., improvements to both inputs and outputs). The first output, however, is emphasized less than the other outputs. The sliders work to choose the direction and thereby steer the projection of the analyzed firm onto the efficient frontier. The figures that indicate direction in IBEN are percentages, and the idea is that they are percentages of the present values for the analyzed firm. Therefore, the correspondence between the IBEN illustration and our framework here is as follows:

$$d = (100\%3.42, 100\%2.30, 50\%1362, 100\%149) = (3.42, 2.3, 681, 149)$$

The resulting benchmark is also shown in IBEN. In our notation, the natural benchmark would be

$$\text{Benchmark} = (x - ed_x, y + ed_y)$$

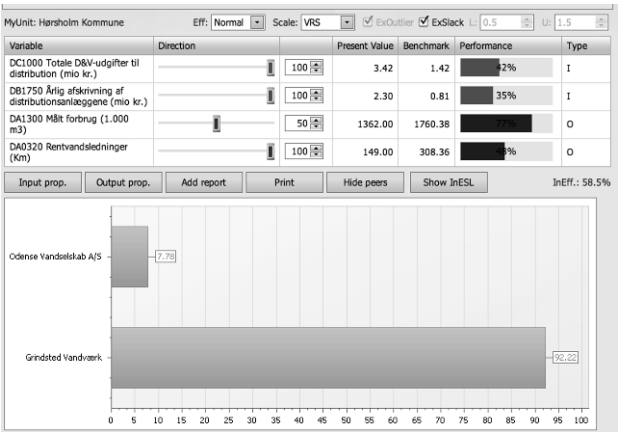


Fig. 2.8 IBEN evaluation of Danish waterwork

but this is not exactly the same as what is shown in the illustration, except with regard to Opex. The reason for this difference is that this benchmark may contain slack and that the slack has been eliminated from the example, cf. also the ExSlack checkbox.

Lastly, we note that IBEN shows the excess value e as InEff and the individual inefficiencies on the different dimensions. Thus, for example, a value of 35% on the Capex line shows that it is only necessary to use 35% of the present Capex level (i.e., 0.81/2.30=35%). In addition, IBEN provides information about the peer units on the lower part of the screen. In this way, the user can see which entities to learn from and how this depends on the direction chosen. IBEN also allows the user to easily remove peers and hereby to re-estimate the technology and directional efficiency on a modified technology.

2.6 Efficiency measures with prices

So far, we have focused on performance evaluations in contexts with a minimum of information. We have assumed that we have firms transforming multiple inputs $x \in \mathbb{R}^m_+$ into multiple outputs $y \in \mathbb{R}^n_+$ using one of several possible production plans T. In addition, we have assumed that we prefer more outputs and fewer inputs. Except for this assumption, we have made no assumptions about the relative importance of inputs and outputs.

In some situations, however, we know a priori the relative weights, prices or priorities that can naturally be assigned to the different inputs and/or outputs. Such information allows us to make more focused evaluations. Moreover, it allows us to decompose efficiency into technical efficiency, associated with the use of optimal

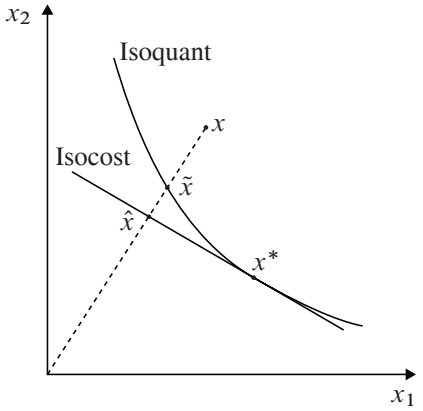


Fig. 2.9 Cost minimum

procedures, and allocative efficiency, associated with the use of optimal combinations of inputs or the production of optimal combination of outputs.

2.6.1 Cost and input allocative efficiency

Let us consider an example in which we have prices associated with the inputs. Let w be the n–the vector of input prices, $w \in \mathbb{R}^n_+$.

In this situation, we can calculate the costs wx of a given production plan (x, y), and thereby evaluate the production plan (x, y) via the cost output combination (c, y), where $c = wx$. In principle, we can conduct efficiency analyses of this, more aggregated, model just as we did with the (x, y) model.

It is now intuitively clear that it is easier to be efficient using the (x, y) model than the (c, y) model because in the latter situation, the firm is responsible not only for picking a technically efficient point on T^E but also for picking the right one to minimize the costs. We shall refer to the latter as the allocation problem and the associated efficiency as *allocative efficiency*.

To formalize this idea, let us assume that a firm has used inputs x, as illustrated in Fig. 2.9.

Ignoring the price information, we can measure Farrell efficiency in the usual way. To distinguish it from other forms of efficiency here, we will now call this the *technical input efficiency* of observation x. As we have seen, it is the maximal contraction of the input bundle and can be calculated as

$$TE = \frac{\|\tilde{x}\|}{\|x\|}$$

where $\tilde{x}$ is the point on the isoquant obtained via proportional scaling for the observed x along the dashed line in the figure.

In the same way, we can measure *cost–efficiency* CE as the ratio between the minimal cost and the actual cost

$$CE = \frac{wx^*}{wx}.$$

The optimal minimal cost input combination x^* is found by solving the cost minimization problem

$$\min\ w'x \quad \text{subject to} \quad (x', y) \in T.$$

The solution to this optimization problem is well known to be the point x^* where the isocost line is tangent to the isoquant as shown in Fig. 2.9.

Cost–efficiency CE is actually also Farrell efficiency in the more aggregate model that uses costs as inputs.

Before we proceed, let us rewrite technical efficiency, TE. It is clear that technical efficiency is also equal to the cost of $\tilde{x}$ compared to the cost of x because the two vectors are proportional. That is, because $\tilde{x} = TEx$, we also have $w\tilde{x} = TEwx$, and therefore

$$TE = \frac{w\tilde{x}}{wx}$$

If we can save 20% of all inputs from x to $\tilde{x}$, we can also save 20% in costs.

Now compare the costs of $\tilde{x}$ and x^*. The difference is the cost of having picked the technically efficient plan $\tilde{x}$ rather than another and less expensive input mix x^*. Thus, the difference represents an allocation problem, and we define *allocative efficiency* as

$$AE = \frac{wx^*}{w\tilde{x}}$$

We see that $AE \leq 1$. If, for example, AE is 0.8, it means that we could have saved 20% by better allocating our funds toward a less expensive but sufficient input mix.

In summary, we now have three different efficiency measures: technical efficiency TE, cost efficiency CE and allocative efficiency AE. The relationship between them is easy to derive:

$$CE = \frac{wx^*}{wx} = \frac{wx^*}{w\tilde{x}}\frac{w\tilde{x}}{wx} = AE \cdot TE$$

This decomposition emphasizes our initial intuition. *To be cost–efficient, the firm must be able to choose the right mix of inputs and use them in a technically efficient manner*. It must use the right resources, and it must use them in the right way.

In closing, we note that if we define $\hat{x}$ as the point on the dotted contraction curve that has the same costs at x^*, then we can also look at the relationship between CE, AE and TE by comparing the length of x, $\tilde{x}$ and $\hat{x}$ as follows:

$$TE = \frac{\|\tilde{x}\|}{\|x\|}, AE = \frac{\|\hat{x}\|}{\|\tilde{x}\|}, CE = \frac{\|\hat{x}\|}{\|x\|}$$

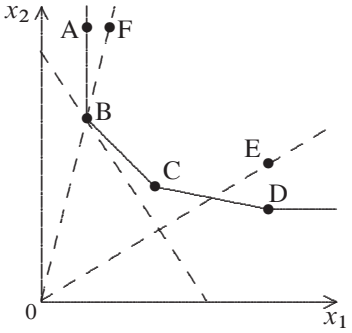


Fig. 2.10 Technology for cost minimization example

i.e., by comparing the lengths of vectors on the dotted line. We see that all of these efficiency measures are smaller than or equal to 1.

Numerical example

Consider a simple example in which six firms A – F have used two inputs to produce one output. The data are provided in Table 2.2.

Table 2.2 Data for cost minimization

Firm	x_1	x_2	y	Costs wx
A	2	12	1	15.0
B	2	8	1	11.0
C	5	5	1	12.5
D	10	4	1	19.0
E	10	6	1	21.0
F	3	12	1	16.5
Price w	1.5	1.0		

We see that all firms have produced the same output, so we can safely look at the problem in the input space. Assuming free disposability and convexity (i.e., that we can produce at least the same outputs with more inputs and that weighted averages (line segments between observations) are feasible as well) we can construct a technology from these observations. The input isoquant (for $y = 1$) of this technology is illustrated in Fig. 2.10 below. The assumptions of free disposability and convexity will be discussed in detail in the next chapter.

The resulting efficiency values are shown in Table 2.3. We see that all firms except E and F are on the frontier and thus are technically efficient; i.e., they have $TE = 1$. The technical efficiency of firms E and F can be calculated by first noting

Table 2.3 Economic efficiency

Firm	*CE*	*TE*	*AE*
A	0.73	1.00	0.73
B	1.00	1.00	1.00
C	0.88	1.00	0.88
D	0.58	1.00	0.58
E	0.52	0.75	0.70
F	0.67	0.67	1.00

that they are projected onto $0.5C + 0.5D = (7.5, 4.5)$ and $B = (2, 8)$ respectively. Thus, for example, the *TE* of F is $2/3 = 8/12 = 0.66$.

Although most of the firms are technically efficient, they have not been equally good at selecting the cost-minimal input mix. These differences become clear when we calculate costs in the last column of Table 2.2. We see that the firm with the lowest costs is B, with a cost value of 11. This result is not surprising given Fig. 2.10, in which the isocost curve is tangent to the isoquant at B. Calculating cost efficiency is now also straightforward. Thus, for example, the cost efficiency for firm A is $CE = 11/15$ because cost efficiency is the minimal cost compared to the actual cost. It is similar to the technical efficiency measure except that we make the evaluation using a one-input (costs) framework.

Lastly, having calculated both *TE* and *CE*, we can easily determine allocative efficiency, defined as $AE = CE/TE$.

We also note that F is allocatively efficient but not technically efficient. This is the case because F is projected onto the cost minimal production plan B when we remove technical efficiency. The classical approach to allocative efficiency that we have introduced here requires one always to measure allocative efficiency at the frontier.

2.6.2 Revenue and output allocative efficiency

A parallel treatment of allocative issues is possible on the output side. Here we look at whether the output mix is optimal in terms of maximizing revenue for a given input. This depends on the output prices $p \in \mathbb{R}^m_+$. An illustration is provided in Fig. 2.11.

As with cost efficiency, we can define revenue efficiency as

$$RE = \frac{py^*}{py}$$

where y is the observed output and y^* the optimal revenue output i.e., the solution to the revenue-optimizing problem

$$\max \; py' \quad \text{subject to} \quad (x, y') \in T.$$

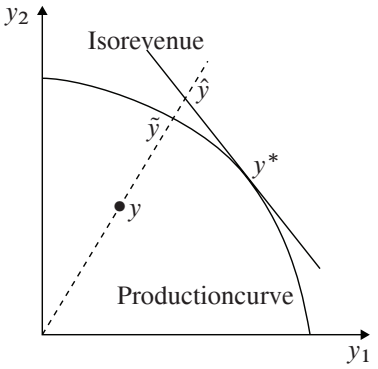


Fig. 2.11 Revenue maximum

We can now rewrite revenue efficiency as

$$RE = \frac{py^*}{py} = \frac{py^*}{p\tilde{y}} \frac{p\tilde{y}}{py} = \frac{py^*}{p\tilde{y}} F = AF \cdot F.$$

Here $\tilde{y} = Fy$, and F is the Farrell output technical efficiency. Therefore, $\tilde{y}$ is the technically efficient point that we obtain when we expand y radially along the dotted line. Also note that we have introduced a shorthand AF for output-oriented allocative efficiency:

$$AF = \frac{py^*}{p\tilde{y}}.$$

Output allocative efficiency is the revenue obtained by choosing the best mix of output relative to the revenue from simply being technically efficient.

To be fully revenue-efficient, a firm must demonstrate both full output technical efficiency and full output allocative efficiency. It must use the best procedures to get the most out of its resources, and it must produce the right mix of services. This concept is sometimes summarized by saying that *it is not enough to do things right; one must also do the right things.*

As in the analyses of the input side, we can also look at this decomposition of revenue efficiency in terms of vector lengths. To see this, let us define $\hat{y}$ as the point on the dotted expansion line that has the same revenue as y^*. We then have

$$F = \frac{\|\tilde{y}\|}{\|y\|}, AF = \frac{\|\hat{y}\|}{\|\tilde{y}\|}, RE = \frac{\|\hat{y}\|}{\|y\|}$$

i.e., by comparing the lengths of vectors on the dotted line, we can calculate all three efficiency values.

2.6.3 *Profit efficiency*

If we have prices w and p on both the input and the output side, we can, of course, also evaluate the firms ability to generate profit and use this as the benchmarking focus. In such situations, we will naturally define *profit efficiency* as

$$PE = \frac{py - wx}{py^* - wx^*}$$

where (x, y) is the observed production plan and (x^*, y^*) is the profit- maximizing production plan, i.e., the solution to

$$\max \; py' - wx' \quad \text{subject to} \quad (x', y') \in T.$$

A small value of PE would be an indication that large profit potentials have been foregone.

Again, one can decompose the inefficiency into different parts related to 1) technical inefficiency, 2) input allocative efficiency and 3) output allocative efficiency. All of these different forms of efficiency describe the firms ability to get the most out of given resources, select a cost-minimal input mix, and select a revenue-maximizing output mix. The decomposition will be somewhat arbitrary depending on the order in which we identify the elements and particularly on the choice of an input- or output-oriented technical efficiency measure. We will not discuss the alternatives in any more detail here.

2.7 Dynamic efficiency

Over time, the behavior and performance of firms are likely to change. We need measures that capture such changes. In addition, the technology is likely to change due to technical progress. These changes make it relevant to measure not only how firms change over time but also how many of these changes are caused by general technological progress and how many can be attributed to special initiatives on the part of individual firms that improve relative to the existing technology.

An example of these dynamic issues is provided in Fig. 2.12 below. We depict the state of one firm during two periods: first, period s and then period t. Likewise, we have two technologies that are relevant for the two periods. We see that the firm has improved in the sense that from s to t, it has moved closer to the s technology. On the other hand, the technology has also shifted, which has made it less costly to produce. Therefore, the firm has not improved as much as we would expect from a general technological development perspective. In period t, it has more excess costs than in period s.

In the benchmarking literature, the most popular approach to dynamic evaluations is the Malmquist index. It works without prices to aggregate the different inputs and outputs.

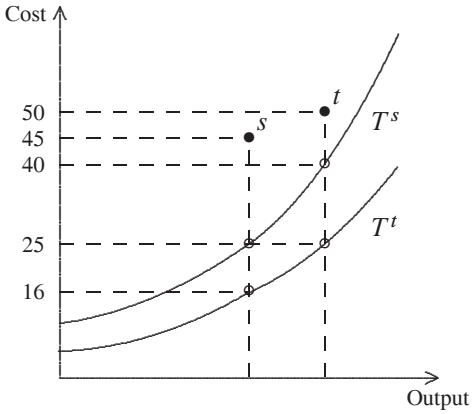


Fig. 2.12 Dynamic change of performance and technology

To explain the idea of the Malmquist index, let $E^i(s,t)$ be a *measure of the performance of firm i in period s against the technology in period t*. It might, for example, be Farrell input efficiency, i.e., $E^i(s,t) = \min\{ E > 0 \mid (Ex^{is}, y^{is}) \in T^t \}$. However, it could also be other measures, including output-based ones, as long as larger values reflect better performance (closeness to the frontier). Note that we now distinguish the technology and the production data depending on the period from which they stem. In this paragraph, we will focus on the performance of firm i, and therefore, we can simplify the notation and use $E(s,t)$ instead of $E^i(s,t)$.

To measure the improvement in firm i from period s to period t, we can look at the changes in efficiency compared to a fixed technology. If we use time s technology as our benchmark, we can look at

$$M^s = \frac{E(t,s)}{E(s,s)}.$$

If the firm has improved from period s to t, $E(t,s) \geq E(s,s)$, and therefore, M^s is larger than 1. If, on the other hand, the firm is less efficient in period t than period s, $E(t,s) \leq E(s,s)$, the ratio is smaller than 1. Therefore, M^s is larger than 1 when the firm improves over time and smaller than 1 if it moves away from the frontier over time. For example, if a firm is 40% efficient in period s and 80% efficient in period t, then it has improved by a factor of 2. This is a natural way to measure the change. If the firm is always producing the same outputs and if it is using the same input mix in the two periods, then in the example, it must have halved its use of resources in period t compared to period s to show this kind of change in its efficiency score. Of course, real problems are more complicated because the input mix and output mix are likely to change as well, but the interpretation of the ratio is still basically the same.

M^s measure the improvement relative to technology s. We might alternatively have used technology at time t as the fixed technology, in which case we would

then get

$$M^t = \frac{E(t,t)}{E(s,t)}$$

Because there is no reason to prefer one to the other, the Malmquist index is simply the geometric average of the two:

$$M(s,t) = \sqrt{M^s\, M^t} = \sqrt{\frac{E(t,s)}{E(s,s)}\,\frac{E(t,t)}{E(s,t)}}.$$

The Malmquist index measures how much a firm has improved from one period s to the next t. The change in performance may, however, be due to two reinforcing or counteracting factors: the general technological progress (or regression) that we would expect everyone to be affected by and special initiatives in the firm that have enabled it to change its performance relative to that of the other firms. We can decompose the Malmquist measure in these two effects by rewriting M as follows:

$$M(s,t) = \sqrt{\frac{E(t,s)}{E(t,t)}\,\frac{E(s,s)}{E(s,t)}}\,\frac{E(t,t)}{E(s,s)} = TC(s,t)\,EC(s,t)$$

where

$$TC(s,t) = \text{ technical change } = \sqrt{\frac{E(t,s)}{E(t,t)}\,\frac{E(s,s)}{E(s,t)}}$$

$$EC(s,t) = \text{ efficiency change } = \frac{E(t,t)}{E(s,s)}$$

The *technical change index, TC* is the geometric mean of two ratios. In both, we fix the firm's production plan at time t and use this as the fixed point against which we measure changes in the technology. If the technology has progressed, we will have $E(t,s) > E(t,t)$ because the technology has moved further away from the given observation (i.e., the first ratio in the geometric mean is > 1). The idea of the second ratio is the same; here we just use the time s version of our firm as the fixed point when we look at technological developments. In summary, the TC measures technological change, and values above 1 represent technological progress in the sense that more can be produced using fewer resources.

The other factor is the *efficiency change index EC*, which measures the *catch-up* relative to the present technology. We always measure this factor against the present technology, asking if the firm has moved closer to the frontier. If so, $E(t,t) > E(s,s)$, and the ratio is larger than 1.

The Malmquist measure is useful to us in understanding how benchmarking results change over time. A firm that has made improvements over the course of a year may be frustrated to learn that it is actually coming out worse in a new benchmarking analysis. The point is, however, that it is not sufficient for a firm to improve compared to itself. The firm must also improve relative to others, and they have also

benefited from general technological progress. Thus, the only way to improve is to catch up to the best, i.e., to get closer to the frontier.

The Malmquist measure and its decomposition are useful in capturing dynamic developments from one period to the next. One should be careful in interpreting results from several periods. One cannot simply accumulate the changes because the index does not satisfy what is called the *circular test*; i.e., we may not have $M(1,2) \cdot M(2,3) = M(1,3)$ unless the technical change is particularly well-behaved (Hicks-neutral). This drawback is shared by many other indices and can be remedied by, for example, using a fixed-base technology.

Lastly, let us mention that some of the ideas in the Malmquist approach can also be used to determine the effects of other changes besides time. We could, for example, let s and t represent two different ways to organize production, two different countries, or two technologies, one with and one without advanced automation (robots). The technological change (TC) in such situations would then reflect the general impact of the technological opportunities created by using alternative organizational forms, operating in one or another country or introducing the use of robots.

Numerical example

To give an example of how the formula is used, let us calculate M, TC and EC in the example shown in Fig. 2.12. Using Farrell input efficiency and observing that the inputs in the example are shown on the vertical axis, we can observe the following directly from the graphs:

$$\underset{\text{(Malmquist index)}}{M(s,t)} = \sqrt{\frac{E(t,s)}{E(s,s)}\frac{E(t,t)}{E(s,t)}} = \sqrt{\frac{40/50}{25/45}\frac{25/50}{16/45}} = \sqrt{\frac{81}{40}} = 1.423$$

$$\underset{\text{(Technical change)}}{TC(s,t)} = \sqrt{\frac{E(t,s)}{E(t,t)}\frac{E(s,s)}{E(s,t)}} = \sqrt{\frac{40/50}{25/50}\frac{25/45}{16/45}} = \sqrt{\frac{5}{2}} = 1.581$$

$$\underset{\text{(Efficiency change)}}{EC(s,t)} = \frac{E(t,t)}{E(s,s)} = \frac{25/50}{25/45} = 0.9$$

This illustrates what can also be inferred from the graph: the firm has improved from period s to t. If we fix the technology, we see that it has moved much closer to the minimal cost curve. The Malmquist index suggests a 42.3% improvement. What is also clear, however, is that this improvement should be expected simply on the basis of the technological developments. In fact, the frontier shift generates a 58.1% improvement cost. So, the firm has not quite been able to follow the trend of technological development but has instead fallen back an additional 10%. The EC and TC effects are multiplicative, such that $EC \cdot TC = M$.

Practical application: Regulation of electricity networks

Most European distribution companies, DSOs, are regulated by competition authorities, cf. Sect. 1.1.3. The single most widely used type of regulation is the revenue-cap regulation, in which the regulator defines ex ante the maximal allowed price companies can change their consumers over the next 3-5 years. A typical scheme would be

$$R^i(t) = C^i(0)Q(0,t)P(0,t)(1-x-x^i)^t, \qquad t = 1,\ldots,5$$

where $R^i(t)$ is the allowed revenue in period t for firm i, $C^i(0)$ is the actual cost of running the DSO in period 0, $Q(0,t)$ is a quantity index reflecting the increase in services from time 0 to t, $P(0,t)$ is a similar index reflecting changes in prices (inflation), and x is a general requirement imposed on all firms and x^i is a specific, additional revenue reduction requirement imposed on DSO i. Hence, the idea is that the regulator allows the DSO to cover its costs but, on a yearly basis, requires it to conduct a general cost reduction of x (e.g., 1.5%) and a specific cost reduction of x^i (e.g., 3%). The advantage of this scheme is that it allows firms to keep what they gain by cutting costs (at least beyond the $x + x^i$ requirement), thus providing them with proper incentives. Also the scheme protects consumers against excessive changes by ex ante requiring charges to fall (with $x + x^i$).

In the implementation of these schemes, a major issue is now how to determine general and individual requirements, x and x^i, respectively. In most cases, solving this problem requires the use of advanced benchmarking. Indeed, x is often established as the frontier shift in Malmquist analyses run on data from a period of some 3-5 years prior to the regulation. Thus, if $TC = 1.02$, the regulator will set $x = 2\%$. Likewise, the setting of x^i is typically informed by a benchmarking model covering, for example, the period $t = 0$ or $t = -1$. The typical benchmarking study will calculate the cost efficiency of each firm and then decide how many years the firm should have to eliminate its incumbent inefficiency, i.e., how quickly it must catch up to best practice. Thus, for example, if a firm has cost efficiency of $CE = 0.80$, it might be asked to partially eliminate this advantage during the regulation period via an extra yearly reduction in costs of, for example, $x^i = 3\%$.

Similar schemes are used to regulate many other sectors as well as to guide budget allocation in public and private organizations.

2.8 Structural and network efficiency

Most of the benchmarking literature is concerned with evaluating the performance of individual firms, i.e., the unit of analysis is firms. It is, however, also possible to evaluate the efficiency of a collection of firms and thus to evaluate if we have the best possible industry structure or if it would pay to move production around, perhaps merging some of the firms and splitting up others. We will briefly illustrate how

such analyses can be conducted and return to more comprehensive and complicated cases in later chapters.

First, consider the possible impact of merging firms 1 and 2, which have used similar inputs to produce similar outputs (i.e., a horizontal merger). Let their present production be (x^1, y^1) and (x^2, y^2), respectively. We do not require that they use exactly the same input and output types because we can always allow the value of some of the dimensions of the x and y vectors to be 0.

If the two units become integrated but continue to operate as two independent entities, they will transform the vector of inputs $x^1 + x^2$ into the vector of outputs $y^1 + y^2$. To evaluate the potential efficiency gains from the merger, we can therefore evaluate the efficiency of the latter transformation, i.e., the use of $x^1 + x^2$ to produce $y^1 + y^2$.

Using a Farrell input approach provides us with the following measure of the potential gains from merging firms 1 and 2:

$$E^{1+2} = \min\{E \in \mathbb{R}_+ \mid (E(x^1 + x^2), y^1 + y^2) \in T\}.$$

Here E^{1+2} is the maximal proportional reduction in the aggregated inputs $x^1 + x^2$ that allows the production of the aggregated output $y^1 + y^2$.

If $E^{1+2} < 1$, we can save via a merger. If $E^{1+2} > 1$, the merger is costly. A score of $E^{1+2} = 0.8$ would suggest that 20% of all inputs could be saved by integrating firms 1 and 2. Likewise, a score of $E^{1+2} = 1.3$ would suggest that integration would necessitate 30% more of all resources. We shall investigate such measures and conduct some useful decompositions in more detail in Chap. 9.

Practical application: Merger control in health care

The evaluation of potential gains from mergers is used in Dutch regulations to shape the health authorities view of proposed mergers. If two hospitals merge, the competition in the sector decreases, and this will generally decrease the quality of care. Industrial economics models of imperfect competition are used to quantify the likely negative market effects. On the other hand, a merger may also be sufficiently efficiency-enhancing and cost-reducing to be attractive despite the reduced competition. To quantify the possible efficiency gains, the Dutch health authority has estimated models of hospital production and set up evaluations of gains like E^{1+2} above. If E^{1+2} is sufficiently small, this will sway the evaluators in favor of allowing the merger.

Rather than merging two or more firms, which may be costly—especially if the technology shows decreasing returns to scale—we can also try to preserve the existing number of firms and simply reallocate production between them. The potential gains from this step can be calculated in the following way. Imagine that we have 3 firms. Generalizations to more firms are straightforward. Let the firms be denoted $k = 1, 2, 3$, and let their original productions be (x^k, y^k), $k = 1, 2, 3$. Assume that we pick new production plans (x^{*k}, y^{*k}) for each $k = 1, 2, 3$ such

that total inputs and outputs stay feasible; i.e., we do not use more aggregated input, $x^{*1} + x^{*2} + x^{*3} \leq x^1 + x^2 + x^3$, and we produce at least the same aggregated output, $y^{*1} + y^{*2} + y^{*3} \geq y^1 + y^2 + y^3$. All of the new production plans must be feasible $(x^{*k}, y^{*k}) \in T$ for all $k = 1, 2, 3$. The largest proportional savings on original input usage that we can achieve via such reallocation can be calculated by solving the following program:

$$\begin{aligned} \min_{H,(x^{*j},y^{*j}),j=1,2,3} \quad & H \\ \text{s.t.} \quad & H(x^1 + x^2 + x^3) \geq (x^{*1} + x^{*2} + x^{*3}), \\ & (y^1 + y^2 + y^3) \leq (y^{*1} + y^{*2} + y^{*3}), \\ & (x^{*j}, y^{*j}) \in T, j = 1, 2, 3. \end{aligned}$$

Therefore, if $H = 0.9$, this means that we can save 10% of all resources used in the three firms by simply moving production around to take advantage of best practices, economics of scale, and economies of scope. So far, we have not made any assumptions about the underlying technology set, T, but if we assume that it is convex, we can actually show that the saving factor H shown above can also be calculated via as simple Farrell input efficiency evaluation of the average firm

$$H = \min_H \{ H \mid (H\frac{1}{3}(x^1 + x^2 + x^3), \frac{1}{3}(y^1 + y^2 + y^3)) \in T \}.$$

Thus, to calculate H, we can simply form the average firm, i.e. a hypothetical firm using the average of all input vectors to produce the average of all output vectors. The Farrell efficiency of this entity is a measure of what can be gained by everyones adjusting to best practices and by reallocating production between the 3 firms. We shall investigate such programs and some useful variations in Chap. 9.

Numerical example

As an example of the reallocation issue, consider a case in which 3 firms have produced 1 output using 1 inputs. The production frontier is given by

$$y = \sqrt{x - 5} \text{ for } x \leq 5.$$

The observed input–output combinations are

$$\begin{aligned} (10, \sqrt{10 - 5}) &= (10, 2.23), \\ (20, \sqrt{20 - 5}) &= (20, 3.87), \\ (30, \sqrt{30 - 5}) &= (30, 5). \end{aligned}$$

We see that they all operate on the efficient frontier, i.e., on an individual basis they cannot improve. However, if they collaborate and share resources and obliga-

tions, they may be able to conserve some of their aggregated input and still produce the same aggregated output. Specifically, following the guidelines above, we can measure the Farrell efficiency of the average firm. The average firm has used (10+20+30)/3= 20 input units to produce (2.23+3.87+5)/3=3.70 output units. The minimal input necessary to produce output of 3.70 is $3.70^2 + 5 = 18.71$. The minimal share of the average input that suffices to produce the average output is therefore

$$H = \frac{18.71}{20} = 0.94.$$

This result shows that via reallocation, this small industry could save 6 % of input. The reason is quite obvious in this simple single–input, single–output case because there are disadvantages to being small based on fixed costs and disadvantages of being large because of diminishing retursn to scale. Therefore, it is more advantageous to operate average-size firms.

2.9 Choice between efficiency measures

The question naturally arises as to which of the many possible efficiency measures to choose. There are several both applied and theoretical aspects of this.

One very important aspect is *controllability*. The inputs and outputs that can be controlled by the entities to be evaluated are important because it is generally not very informative or motivating to be judged on the basis of factors that you cannot control. Therefore, the choice between input- and output-based evaluations, between general evaluations or conditional evaluations where some factors are fixed, and between allocative and technical efficiency depends very much on controllability.

The time perspective is relevant because in the long run, more factors are usually variable. The level in a hierarchy that is evaluated is relevant. A divisional manager may, for example, be evaluated for technical efficiency, while an officer at the headquarters who is responsible for resource allocation may be more correctly evaluated based on allocative efficiency or, if prices are not available, using structural efficiency measures. A hospital may not have much control over demand, and as a result, input-based evaluations may be more relevant, while a farmer may have many fixed resources (land, etc.) and, therefore, should be evaluated more in terms of the output.

More generally, the *intended use* of the efficiency score is crucial. In a learning experience, the exact efficiency measurement is less important than the ability to find relevant peers taking into account the firms own preference, strategies, etc. The directional distance function approach may be particularly useful here due to its flexibility. In an allocation application, the distinction between fixed and variable inputs and outputs is often important, which might lead us to favor a Farrell approach, with some inputs and outputs that are non-discretionary (or even a directional distance function approach). In an incentive application, the task is to find an aggregation of performance that allows optimal contracting. We will see in later

chapters that one can actually provide incentive rationales for radial measures like the Farrell approach.

On a very specific level, *ease of interpretation* is also important. One of the advantages of the Farrell measure in applications is that it is very easy to interpret. One can come up with many more or less ingenious ranking systems, and those that do not perform well may have very strong objections as to how the ranking was constructed and how the different performance dimensions were aggregated and weighted. One important element of the Farrell measure, however, is that it does not weigh the different dimensions. If a firm is not performing well according to this measure, it is very difficult for that firm to explain away the results because it is underperforming in all areas rather than just in one potentially overrated dimension. This is because the Farrell measure uses proportional changes. This argument can actually be given a game theoretical formalization, as we will show in Chap. 5.

As a last practical concern, let us mention *data availability* and *computations ease*. The more we know about values (prices, priorities), the more focused the evaluations can become. Prices for inputs, for example, enable us to conduct cost efficiency analyses that decompose efficiency into allocative and technical efficiency, which will provide us with more information han a pure technical efficiency analysis would. Likewise, using data from several years allows more robust evaluations and may possibly allow us to separately consider general productivity shifts and catch-up effects. Additionally, in more advanced applications involving, for example, complicated structural and network models, computational issues shall be considered. It is less interesting to dream up complicated calculations if they are very difficult to implement because the resulting programs become too non-linear, for example.

From a more theoretical perspective, we may compare the general properties of different measures using *axiomatic theory*. Some key results are given in Sect. 2.12. As emphasize there, the Farrell measure has several advantages but suffers from one problem: a lack of what is called indication. A firm may be efficient in the Farrell sense even if it is in fact not fully (Koopmans) efficient.

It is also important to keep the rational ideal model in mind when considering indices of technical efficiency. Ideally, efficiency should reflect utility effectiveness because efficiency is a sort of proxy for *utility effectiveness*. We know that dominance relationships are maintained under utility effectiveness in the sense that if one firm dominates another, then it is also more utility effective. We cannot, however, be sure that inefficient firms are less utility-effective then some efficient ones. Therefore, although efficiency provides a useful filter, efficiency is not a sufficient condition for firm effectiveness, and one should not be too fixated on the ability to make efficiency evaluations based on a minimum of assumptions. It is still important to think of ways to elicit preferences and make evaluations that more closely capture our preferences. After all, small improvements of the right type may be more valuable than large improvements to less important aspects.

2.10 Summary

In this chapter, we have taken a somewhat closer look at the general problem of evaluating and quantifying the performance of a firm by gauging it against a technology. We have defined efficiency as using the least resources to produce the most services, and we have looked at different ways to measure efficiency levels. We have covered the most widely used measure, the Farrell efficiency measure focusing on proportional improvements to inputs or outputs, and we have discussed alternative approaches like directional distance functions with excess, an additive measure of the number of times a given improvement bundle is feasible. We have also discussed how preference or price information allows more informative evaluations, including decompositions spotlighting allocative and technical efficiency factors. We have shown how one can distinguish between frontier shifts and catching up in a dynamic context and how structural efficiency can be evaluated by looking at networks of firms. Lastly, we have discussed some key concerns related to the choice between alternative measures. Some more advanced material, including the axiomatic characterization of some classical measures, is provided in Sect. 2.12 below.

2.11 Bibliographic notes

The notion of efficiency is used throughout economics and is perhaps most well-known in the context of the Pareto efficiency concept, wherein the outcomes for several individuals are compared using the efficiency criterion. A solution Pareto dominates another if, and only if, it makes someone better off without making anyone worse off. In multiple criteria decision-making, a main theme is how to find and choose among efficient alternatives, c.f. e.g., Bogetoft and Pruzan (1991). In a production economics context, the traditional reference is Koopmans (1951). The idea behind all related concepts is the same, however: we avoid weighing different persons, different criteria or different inputs and outputs together by using a more is better than less approach and looking for improvements that occur in some area without creating worse performance in others. In Bogetoft and Pruzan (1991), appendix 1, we formalize how efficiency is related to the rational ideal evaluations that economists seek to make.

The focus on proportional improvements was suggested by Debreu (1951) and Farrell (1957). The inverse of Farrell, the Shephard distance function, is due to Shephard (1953, 1970). The use of discretionary and non-discretionary dimensions is described in many textbooks: for example, Charnes et al (1995). However, this use dates back at least to Banker and Morey (1986).

The graph hyperbolic efficiency measure was suggested in Färe et al (1985), while basic work on the excess function was done by Luenberger (1992) and Chambers et al (1998). The idea of constructing interactive benchmarking systems was suggested in Bogetoft and Nielsen (2005) and Bogetoft et al (2006a) and commer-

cialized in the Interactive Benchmarking IB^{TM} software from www.ibensoft.com used by Danish Waterworks.

The idea of allocative efficiency dates back to at least Debreu (1951) and Farrell (1957), while the Malmquist index dates back to Malmquist (1953) and was made popular by Caves et al (1982) and Färe et al (1994). There is a large body of literature on alternative modes of decomposition. Bogetoft et al (2006b) provides an alternative definition of allocative efficiency that allows us to calculate allocative efficiency without assuming that technical efficiency has first been eliminated.

The idea of structural efficiency dates back to at least Farrell (1957) on p.262. He defined structural efficiency as "the extent to which an industry keeps up with the performance of its own best firms" and suggested that it can be measured by comparing the horizontal aggregation of the industry's firms with the frontier constructed from its individual firms. A related approach is the average unit approach suggested by Førsund and Hjalmarsson (1979). In a recent study, Andersen and Bogetoft (2007) developed a DEA-based reallocation model to study the potential gains from redistributing fishery quotas among Danish vessels. An interesting result was that the redistribution of production might be just as useful as the learning of best practices. This is relevant because it may be optimistic to suppose that all units can adopt best practices, at least in the short run, and reallocations off the frontier should therefore be considered, cf. also Bogetoft et al (2006b) The idea of interpreting this result as the possible effect of a reallocation program calculating H comes from Bogetoft and Wang (2005). The application for merger control is developed in Bogetoft and Katona (2008), while the application for the reallocation of agricultural production is described in Andersen and Bogetoft (2007) and Bogetoft et al (2009). We discuss structural efficiency and network models in more detail in Chap. 9, where we provide more references.

The link between efficiency and decision theory formalized in the appendix builds directly on Theorem 1 in Bogetoft and Pruzan (1991), where a proof is also provided.The axiomatic approach to efficiency evaluations was initiated by Färe and Lovell (1978). They worked with axioms 2, 3 and 4 below. This was followed by work by Russell (1985, 1987, 1990), Zieschang (1984), and others. Axiomatic characterizations of special directional distance measures and discussions of their relationship to bargaining theory are given in Bogetoft and Hougaard (1999).

2.12 Appendix: More advanced material on efficiency measures

As an appendix to this chapter on efficiency measures, we will now present some more technical material that can be skipped during a first reading.

2.12.1 The rationale of efficiency

It is of course possible to identify more precise and profound motivations for reliance on efficiency. To consider one such motivation, we will now look at efficiency in a decision theoretical context.

The basic economic model of (individual) choice is the *rational ideal model*. The rational ideal model depicts an economic entity (an individual or system) as seeking the best means to his desired ends; it is defined by the set of alternatives available and ones preferences regarding them.

Let us assume that a firm has transformed m inputs $x^* \in \mathbb{R}^m$into n outputs $y^* \in \mathbb{R}^n$. Additionally, let the objective or preference function be given by

$$U : \mathbb{R}^{m+n} \to \mathbb{R}$$

where $U(x, y)$ is the utility attached to a production plan (x, y) Also, let us assume that the set of feasible input-output vectors is

$$T \subseteq R^{m+n}.$$

In this set-up, we have that (x^*, y^*) is *optimal* if and only if it solves the *basic decision problem*

$$\begin{array}{l} \max U(x, y) \\ s.t. \quad (x, y) \in T \end{array} \tag{2.1}$$

i.e., if and only if the firm has made the best, most effective use of its potential.

In practice, this ideal evaluation can seldom be conducted. A common obstacle is that the feasible production plans T are not known. Another is that the firm or the evaluator may not have clear-cut expressions of the aggregate performance evaluation criterion $U(.)$. In public sector contexts, for example, where the agent could be a school, court or police station, it is often hard to imagine explicit delineations of the production options. Additionally, the multiple outputs produced will be difficult to aggregate, such that an explicit preference structure is usually not available.

One perspective on the modern theory of productivity analysis is that it allows us to make evaluations in contexts with incomplete information about options T and preferences $U(.)$. This is done by focusing on efficiency instead of effectiveness and by focusing on efficiency relative to a constructed technology rather than in relation to the underlying true but unknown technology.

Let us now focus on the first problem, the lack of information about $U(x, y)$ and the resulting need to shift our attention from effectiveness to efficiency.

As a matter of notation, recall that an input–output combination $(x', y') \in T \subseteq R^{m+n}$ is efficient relative to the technology T if and only if

$$\forall (x, y) \in T : x \leq x', y \geq y' \Rightarrow x = x', y = y'.$$

The set of *efficient* plans is denoted T^E.

The focus on efficiency is natural. On the one hand, efficiency is not too strong a requirement because one can always find an optimal production plan among the efficient ones, and on the other hand, we cannot strengthen the efficiency requirement because any efficient plan may be the uniquely optimal plan based on one of the underlying but unknown preference functions.

We formalize these reflections in the following proposition.

Proposition of Rational Efficiency. *For a basic decision problem (2.1) where U is weakly increasing in y and weakly decreasing in x, i.e., $x \leq x', y \geq y' \Rightarrow U(x, y) \geq U(x', y')$, we have that*

1. *for any (x^*, y^*) optimal in (2.1), there exists a $(x, y) \in T^E$ such that $U(x, y) = U(x^*, y^*)$, and*
2. *for any $(x^*, y^*) \in T^E$, there exists a U such that (x^*, y^*) is a unique solution to (2.1).*

This is a straightforward modification of a well-known result in decision theory.

According to the proposition, we do not lose anything by focusing on efficient production plans. By the first bullet, an optimal alternative can always be found among the set of efficient alternatives. However, the set that we consider to find the optimal alternative cannot a priori be a smaller set than T^E if all we know about the preference function or the overall evaluation criteria U is that they are weakly increasing. By the second bullet, any efficient alternative may turn out to be the only optimal plan for a weakly increasing U. Thus, *the efficient set T^E is the smallest sufficient set of alternatives to consider.*

As noted in Sect. 2.4 and below in the axiomatic characterization, Farrell efficiency does not guarantee efficiency because there may be slack left when we project a point onto the frontier of the technology. This should not, however, disturb us too much. After all, when we use the radial measures, we simply find more Farrell-efficient points than truly efficient points, i.e. we do not exclude any interesting points a priori, but we may leave uninteresting points in the Farrell efficient set.

2.12.2 Axiomatic characterization of efficiency measures

To understand the pros and cons of different benchmarking approaches, it is useful to develop a basic understanding of the properties of the efficiency measures that we use. Here we introduce some desirable properties of efficiency measures and then record which of them the Farrell measures (and a few other measures) have.

To simplify the exposition, we focus on the input space. The technology can therefore be defined as the input set L, i.e. the set of input combinations $x \in \mathbb{R}^m_+$ that can produce a fixed amount of output $y \in \mathbb{R}^n_+$. Formally, $L(y) = \{ y \in \mathbb{R}^m_+ \mid (x, y) \in T \}$ and to simplify we just write L as y is fixed in what follows. With standard regularity assumptions on T if follows that $L \subset \mathbb{R}^m_+$ has the properties

of being non-empty, closed and free disposable, and every $x \in L$ can produce y. Define the weakly efficient (Farrell efficient) subset of L (i.e., the isoquant of L,) as $I = \{ x \in L \mid \lambda \in [0,1[\Rightarrow \lambda x \notin L \}$, and the efficient (Koopmans efficient) subset of L as $L^E = \left\{ x \in L \middle| \forall x' \in \mathbb{R}^m_+ : x' \leq x \wedge x' \neq x \Rightarrow x' \notin L \right\}$, in two dimensions the part of the isoquant that does not contain vertical or horizontal parts.

To be of any general interest, an efficiency concept must be applicable to a reasonably large class of technologies: for example, any technology in a set $\mathcal{L}$ with the properties listed for L above. Note that an efficiency measure basically maps a production plan and a technology into the real numbers. We can formally define it in the following way. *An efficiency measure or index* is a function

$$\epsilon : \mathbb{R}^m_+ \times \mathcal{L} \to \mathbb{R}$$

such that $\epsilon(x, L) \in [0, 1]$ for $x \in L$.

We see that the Farrell efficiency measure satisfies these conditions.

Another measure that has been around for several years is the *Färe-Lovell efficiency* index.

$$E_{FL}(x, L) = \min\Big\{ \frac{1}{\#\{i \,|\, x_i > 0\}} \sum_{i=1}^{m} \lambda_i \;\Big|\; (\lambda_1 x_1, \ldots, \lambda_m x_m) \in L,\ \lambda_i \in [0,1],\ x \in L \Big\}.$$

The idea of this measure is that we try to minimize the average of the input-specific contraction factors; i.e. we conduct individual contractions of the different inputs. Hence, this process does not necessarily lead to proportional reductions as in the Farrell case. Graphically, and presuming that all $x_i > 0$, $i = 1, \ldots, m$, the measure corresponds to comparing x to the point on L that minimizes a cost function with prices $(x_1^{-1}, x_2^{-1}, \ldots, x_m^{-1})$. The reason is that minimizing $\sum_{i=1}^m \lambda_i$ under the restriction $(\lambda_1 x_1, \ldots, \lambda_m x_m) \in L$ is equivalent to minimizing $\sum_{i=1}^m x_i^{-1} \tilde{x}_i$ over $\tilde{x} \leq x, \tilde{x} \in L$. Simply substitute using $\lambda_i x_i = \tilde{x}_i$.

A third measure combining the two is the *Zieschang index* defined as

$$E_Z(x, L) = E(x, L) E_{FL}(Ex, L)$$

which corresponds to Farrell efficiency multiplied by the Färe-Lovell efficiency of the Farrell projected input combination.

It is easy to see that

$$E(x, L) \geq E_Z(x, L) \geq E_{FL}(x, L).$$

Now let us consider some general and desirable properties:

Commensurability / Invariance to permutations and rescaling (A1) For all $m \times m$ matrices with exactly one non-zero and positive element in each row and column, we have that $\epsilon(x, L) = \epsilon(Ax, AL)$.

Indication (A2) $\epsilon(x, L) = 1$ if and only if $x \in L^E$.

Homogeneity of degree −1 (A3) $\epsilon(\lambda x, L) = \lambda^{-}1\epsilon(x, L)$ for all λ where $\lambda x \in L$.
Monotonicity in inputs (A4) $x' \geq x, x' \neq x$ implies $\epsilon(x, L) > \epsilon(x', L)$.
Continuity (A5) $\epsilon(\lambda x, L_e)$ is a continuous function of $\lambda \geq 1$ when $L_e = \{x \in \mathbb{R}^m_+ \mid x \geq e\}$ and $x \in L_e$.

We interpret these properties as follows:

Commensurability (A1) means that efficiency is not affected by different permutations of inputs, i.e. it does not matter in what order we list the inputs. Moreover, efficiency is independent of linear re-scalings of the different inputs. Thus, for example, it does not matter if we measure in kg or tons. Both E, E_{FL} and E_Z clearly have this property.

Indication (A2) means that we only assign the value 1 to points that are efficient (in the Koopmanns sense). The Farrell measure does not have this property because the radial project may end at a vertical or horizontal part of the isoquant. This is the major drawback of the Farrell measure and one of the motivations for the Färe-Lovell and Zieschang indices.

Homogeneity of degree -1 (A3) means that if we double the inputs, we halve the efficiency. Farrell efficiency and Zieschang efficiency satisfy this, but the Färe-Lovell's index does not.

Monotonicity of inputs (A4) requires that if we increase the usage of at least one input, we lower the efficiency score.

Continuity (A5) requires that if we have a Leontief technology and vary the input consumption proportionally for the technology, then the efficiency score will vary continuously. This is a desirable property in practice because we do not want small data errors to have dramatic impact on the efficiency score. Unfortunately, this requirement is not easy to fulfill.

Although these properties seem reasonable, they are not easy to fulfill. As a general *non-existence theorem*, we note that one cannot construct a measure that satisfies A2, A3 and A4 simultaneously for the ample class of technologies, $\mathcal{L}$, that we have considered here. In this sense there is no best efficiency measure to always be used in efficiency and benchmark analysis.

Table 2.4 summarizes the properties of the Farrell, Färe-Lovell and Zieschang measures. In the table, (Yes) means Yes as long as we only consider strictly positive input vectors.

Table 2.4 Properties of efficiency measures

Property	E	E_{FL}	E_Z
Commensurability (A1)	Yes	Yes	Yes
Indication A2	No	Yes	Yes
Homogeneity A3	Yes	No	Yes
Monotonicity A4	No	(Yes)	No
Continuity A5	(Yes)	No	No

Chapter 3
Production Models and Technology

3.1 Introduction

In Chap. 1, we briefly introduced the concept of a production set or a technology as a way to characterize the production possibilities in a given application. This concept is crucial in advanced benchmarking because it defines the set of possible performance outcomes against which we can evaluate the actual performance of a given firm.

All analyses related to production depend in one way or another on technology. Technology shows how inputs can be turned into outputs, how inputs can be substituted for each other, how outputs depend on inputs, and whether outputs are the result of a joint or a united process.

In this chapter, we discuss the technology set in more detail. We emphasize some common properties of technology sets: disposability, convexity and return to scale. A good understanding and feel for these properties is important in benchmarking because they drive much of the comparison process. In the appendix to this chapter, we also explain the relation between these ideas and related concepts such production correspondences, consumption correspondences and cost functions. We also include a brief introduction to duality. This more advanced material is less important for a first reading.

3.2 Setting

A firm can be thought of as a decision-making unit that chooses a production plan (i.e., a combination of inputs and outputs). From this perspective, a firm serves to transform inputs into outputs. This is illustrated in Figure 3.1.

We assume that we have K firms indexed $k = 1, \ldots, K$. Each firm uses m inputs to produce n outputs. Some of these outputs may be zero; therefore, the setting is not that restricted.

P. Bogetoft and L. Otto, *Benchmarking with DEA, SFA, and R*, International Series in Operations Research & Management Science 157, DOI 10.1007/978-1-4419-7961-2_3,

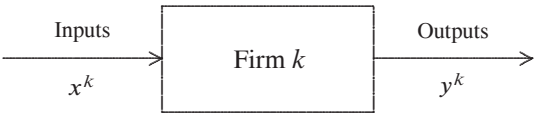


Fig. 3.1 Simple firm

We generally denote the *inputs* for firm k as the m–vector

$$x^k = (x_1^k, \ldots, x_m^k) \in \mathbb{R}_+^m$$

and the *outputs* for firm k as the n–vector

$$y^k = (y_1^k, \ldots, y_n^k) \in \mathbb{R}_+^n.$$

A *production plan* for firm k is thus a pair of input and output vectors

$$(x^k, y^k) \in \mathbb{R}_+^m \times \mathbb{R}_+^n.$$

Note how we use superscripts to denote firms and subscripts to denote the different types of inputs and outputs. When we do not have subscripts, we consider all the inputs or outputs in a vector format.

A final bit of common notation: we use $\mathbb{R}_+ = \{a \in \mathbb{R} \mid a \geq 0\}$ and $\mathbb{R}_{++} = \{a \in \mathbb{R} \mid a > 0\}$. Thus, we presume for now that both inputs and outputs are nonnegative numbers, i.e., that they are positive or zero.

For easy reference, we list the common notation in used in this book in the Acronyms and Symbol sheet.

Our starting point is therefore observations of data from K firms in the form of outputs $(y^1, \ldots, y^K)$ and inputs $(x^1, \ldots, x^K)$. The inputs and corresponding outputs for the different firms can be gathered into a table like Table 3.1.

Table 3.1 Data

Firm	Input	Output
1	x^1	y^1
2	x^2	y^2
$\vdots$	$\vdots$	$\vdots$
K	x^K	y^K

3.3 The technology set

The general idea in benchmarking is that the firms we compare have a common underlying technology as defined by the *technology or production possibility set* T,

$$T = \{\, (x, y) \in \mathbb{R}^m_+ \times \mathbb{R}^n_+ \mid x \text{ can produce } y \,\}.$$

The technology is determined by the social, technical, mechanical, chemical, and biological environment in which the production process takes place.

In many applications, the underlying production possibility set (i.e., the technology) is unknown. It is therefore necessary to estimate the technology set based on observed data points and then to evaluate the observed production of a firm relative to the estimated technology.

Let us assume for now that data are precise and that no random elements are involved in the production. This means that the actual observations must belong to T, i.e.

$$(x^k, y^k) \in T \qquad k = 1, \ldots, K$$

It follows that the *smallest* set that contains data is

$$T = \{(x^1, y^1), \ldots, (x^K, y^K)\}.$$

To prepare for the following, we can express this differently also by saying that T is the set of (x, y) values for which there exists a k such that $(x, y) = (x^k, y^k)$; i.e.

$$T = \left\{\, (x, y) \in \mathbb{R}^m_+ \times \mathbb{R}^n_+ \mid \exists k \in \{1, \ldots, K\} : (x, y) = (x^k, y^k) \,\right\}. \tag{3.1}$$

In general, this is not the most interesting technology. New data will almost certainly enlarge the technology set, and if a firm wants to change its production process within this technology set, it can only do what one of the other firms have already dobe.

At the other extreme is the *largest* possible technology set,

$$T = \mathbb{R}^m_+ \times \mathbb{R}^n_+$$

under which anything can produce anything. This is also not a very interesting model since it is not seldom realistic.

We will now look for a technology set that lies between the smallest possible and largest possible technology sets containing data. We will still use observations as our starting point, but we will add assumptions to enlarge the technology set sufficiently to make it interesting but not enough to allow just anything. Put differently, we will discuss ways to make interpolations and extrapolations of the observations.

Before we do so, however, we should note that there are situations in which the smallest technology set is actually relevant.

Practical application: Bulls

In 2005-2007, a group of Danish economists and quantitative geneticists experimented with the use of benchmarking as breeding support. The data set included more than 1500 Danish bull of a particular breed (SDM) that was described in terms of 14 dimensions, cf. Table 3.2. Each dimension was summarized as an index between 1 and 100, with 100 being the best, and they could therefore be thought of as outputs.

Table 3.2 Bull data

Output indices
Y-index, Total merit, Body, Feet and legs, Mammary system, Milking speed, Temperament, Calving index, Daughter fertility, Mastitis resistance, Birth index, Longivity, Other health traits, Beef production

In this case, we did not have any inputs, although the cost of semen could have been an obvious choice. We also did not use the bull data directly; instead, we used predicted properties of the calves that would be born from a given cow (specified by a user), having mated with each of the 1500+ bulls. In this case, the technology is really the set of these 1500+ expected calves. It would not make sense, for example, to take the average of two calves (unless perhaps one randomized the choice of semen). However, due to the size of the set of potential calves, it is still possible to make interesting comparisons.

3.4 Free disposability of input and output

Our first assumption is that we can dispose of unwanted inputs and outputs. Of course, if prices are positive, we do not want to simply dispose of outputs if we can actually sell them or buy inputs that we do not use, but for now, we only consider the technological possibilities without considering anything that involves markets or preferences.

Thus, the first idea is that if we can produce a certain quantity of outputs with a given quantity of input, then we can also produce the same quantity of outputs with more inputs. One way to interpret this assumption is to say that we can freely dispose of surplus inputs. We call this assumption the *free disposability of input*. We can formalize this idea by saying that if $(x, y) \in T$ and $x' \geq x$, then $(x', y) \in T$, i.e.

$$(x, y) \in T, x' \geq x \Rightarrow (x', y) \in T.$$

Likewise, if a given quantity of inputs can produce a given quantity of outputs, then the same input can also be used to produce less output—we can dispose of surplus output for free. We call this assumption the *free disposability of output* and we can formalize it by saying that if $(x, y) \in T$ and $y' \leq y$, then $(x, y') \in T$, i.e.

$$(x, y) \in T, y' \leq y \Rightarrow (x, y') \in T.$$

When we combine the two assumptions, we derive the assumption of the *free disposability of input and output*: when $(x, y) \in T$, $x' \geq x$, and $y' \leq y$, then $(x', y') \in T$; i.e.

$$(x, y) \in T, x' \geq x, y' \leq y \Rightarrow (x', y') \in T$$

Let us draw a picture of this assumption. One observation ($K = 1$) yields the situation depicted in Fig. 3.2. The vertical dashed line below the observation illustrates

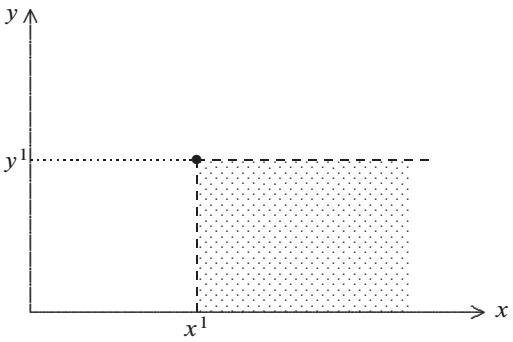


Fig. 3.2 Free disposability: one firm

the production of lower output than y^1 using input x^1, and the horizontal dashed line to the right of the observation illustrates the production of y^1 with more input than x^1. The shaded area indicates the free disposability of both input and output. Thus, based on one observation and this assumption, we have already developed a (simple) technology set that will not necessarily be altered based on a new observation.

When we have more data points like (x^1, y^1), (x^2, y^2), (x^3, y^3) and (x^4, y^4) in Fig. 3.3, the technology set is any input-output combination below and to the right of the data points (i.e., the shaded area in the figure). We see that this set does not really depend on observation (x^4, y^4) because we can infer the feasibility of this point based on the feasibility of (x^3, y^3) and the assumption of free disposability. Therefore, we have a technology that is somewhat more informative that the set of observations.

The technology constructed from a set of observations and the free disposability assumption is often called *the free disposable hull (FHD)* in the benchmarking literature. We can formalize this concept by saying that $(x, y) \in T$ if there is a

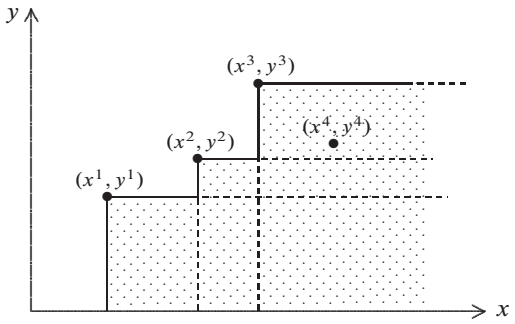


Fig. 3.3 Free disposability

$k \in \{1, \ldots, K\}$ such that $x \geq x^k$ and $y \leq y^k$. Now we can write the technology set T in a manner similar to that indicated in Eq. (3.1) on page 59, noting that an input-output combination(x, y) is feasible if and only if there exists an observation$(x,^k , y^k)$ such that $x \geq x^k$ and $y \leq y^k$; i.e.,

$$T = \left\{ (x, y) \in \mathbb{R}^m_+ \times \mathbb{R}^n_+ \mid \exists k \in \{1, \ldots, K\} : x \geq x^k, y \leq y^k \right\}. \tag{3.2}$$

Free disposability means that inputs and outputs can freely be disposed off—or, to put it differently, *we can always produce fewer outputs with more inputs*. In some instances of joint production, this may not hold. It may not be possible to reduce an unattractive type of output such as CO_2 emissions without a corresponding reduction in attractive outputs like car transportation. Correspondingly, it may not be possible to reduce an output like manure without also reducing the output of pigs. To model such technologies, we use weaker types of disposability assumptions. For example, we may assume that inputs or outputs can be reduced proportionally.

Still, in most cases, free disposability is a safe and weak regularity assumption in the construction of an empirical reference technology. Moreover, this assumption has considerable *appeal in applications* because the peer units can be directly identified and are real units rather than units constructed via some mathematical combination of many units. In Fig. 3.3, for example, it is clear that (x^4, y^4) is doing worse than (x^3, y^3) because it is using more inputs to produce fewer outputs. If the data set is sufficiently large (i.e., if the number of firms K is large relative to the number of inputs m and outputs n needed to describe the activities), then free disposability is also sufficiently powerful to create enough relevant comparisons. If the data set is small, the discriminatory power of the analyses will tend to weaken in the sense that almost all firms will be on the boundary of the constructed technology set and will therefore be efficient, with no opportunity to learn better practices. A few large-scale benchmarking projects based almost entirely on the FDH technology are described in the next section.

Practical application: Credit unions

In 2004-2006, an international team of benchmarking professors was engaged by the Credit Union National Association (CUNA) in Madison, Wisconsin, to develop a benchmarking environment. The set of available data was massive and involved more than 700 variables from more than 10,000 credit unions in each of 6 periods. The system that was designed and implemented, the Credit Union Benchmarker (CUB), is still in use today; it took advantage of the large data sets by basing the technology estimations primarily on the free disposability property.

The aim of the CUB was to generate relevant comparisons by taking into account a multiplicity of inputs and outputs in accordance with the system view. Models for whole credit unions and for different sub-processes were created. The aim was also to give users flexibility in choosing a perspective and to allow users flexibility in their choice of potential peers. More specifically, the user could choose m inputs, n outputs and K credit unions to characterize the technology. A user could thereby combine the power of observed data with subjective beliefs and preferences. For instance, these beliefs and preferences might be reflected in the set K of credit unions that the user found it worthwhile to compare. The credit unions K could be chosen to ensure the use of similar technology (e.g., because industry-specific credit unions are believed to have more similarities than industry-specific and regional credit unions). The credit unions K could also be chosen to reflect preferences (e.g., a preference for learning from credit unions in the same state or region rather than those in a different one). Because of the large number of observations, free disposability was sufficient to generate interesting results in most cases. To support other cases and provide a richer set of information, the system also included the option of invoking restricted rescaling and fuzzy comparisons to enlarge the technology set (see below).

Practical application: Universities

In 2006-2008, a benchmarking system using similar ideas was developed as part of the Aquameth project under the EU Prime Network of Excellence. The focus of the project was on the strategic steering of universities in Europe based on the collection and integration of a large data set covering individual universities in the United Kingdom, Spain, Italy, Norway, Portugal and Switzerland. This was the first dataset available internationally that was based on time series of micro-based indicators at the university level. The dataset included information on some 300 universities, with the UK and Italy as the largest groups. Data from up to 10 years was combined, yielding a total data set of some 2450 university-years and almost 60 variables used as proxy inputs and outputs (see Table 3.3).

In a university context, using multiple years may be important because of the considerable timelag in the production of research outputs.

It is clear from the list of variables that many model specifications are possible. A highly aggregate model focused on both the teaching and the research mission

Table 3.3 University data

Financial Inputs	Physical Inputs	Outputs
Total funds, Student fees, Government funding, EU funding, Private funds, Income from assets, Other funds, Total expenses, Academic staff expenses, Non-academic staff expenses, Other current expenses, Capital expenses, Other expenses	Total academic staff, Full professors, Full professors (female), Full professors (male), Associate professors, Associate professors (female), Associate professors (male), Researchers, Researchers (female), Researchers (male), Other academic staff, Other academic staff (female), Other academic staff (male), Technical staff, Administrative staff, Lecture rooms, Scientific laboratories, Computer labs, Linguistic labs, Libraries, Total surface	Enrolled students, Enrolled students from outside region, Enrolled students (male), Enrolled students (female), Graduate students, Graduate students from outside region, Graduate students (male), Graduate students (female), PhD students, PhD recipients, Masters students, Masters degree recipients, Publications, Patents, Revenue from patents, Spinoff, Cooperation agreements, Co-publications with industry

of a university could involve inputs like academic staff and non-academic staff and outputs like graduate students and number of publications. It is also clear that different users want to focus on different sub-models, and the aim of our contribution is therefore to enable the users to select the mission, as represented by the included inputs m and outputs n, and the relevant set of universities believed to have the technological capacities relevant to this mission, i.e., the K. Despite of the large number of observations, the fact that the panel dataset is not balanced means that in many comparisons (i.e., for many choices for m, n and K), the available dataset is considerably smaller. Still, the basic technology set used to construct technologies is based on the free disposability assumption. As in the case of credit unions, we also allow some rescaling and fuzzy comparisons, which is even more important in the present case because the data set is smaller.

3.5 Convexity

A very powerful property that is often assumed in economics in general and benchmarking in particular is that of convexity. In fact, in economics, convexity is so common that we often take it for granted. In benchmarking, convexity serves the role of enlarging the technology, especially when there are only a few observations available. In turn, convexity also creates technologies that are better able to distinguish between average performance and best practices. We will now explain the idea of convexity and discuss when it is appropriate to introduce this concept and when it is not.

If we have two feasible production plans, it is often assumed that all weighted averages of the two are also feasible. In geometric terms, this would mean that for any two points in the technology set T, the plans on the line between them are also in T. In mathematics, a set T with this property is referred to as convex. A common assumption in benchmarking is can therefore be summarized as

$$T \text{ is convex.}$$

Formally, the set T is *convex* if for any two points $(x^0, y^0) \in T$, $(x^1, y^1) \in T$, and any weight $0 \leq \lambda \leq 1$, the weighted sum $(1-\lambda)(x^0, y^0) + \lambda(x^1, y^1)$ is also in T; i.e.,

$$(x^0, y^0) \in T, (x^1, y^1) \in T, 0 \leq \lambda \leq 1 \Rightarrow (1-\lambda)(x^0, y^0) + \lambda(x^1, y^1) \in T$$

The weighted sum of the two plans

$$(x^\lambda, y^\lambda) = (1-\lambda)(x^0, y^0) + \lambda(x^1, y^1) \qquad (0 \leq \lambda \leq 1)$$

is called a *convex combination* of (x^0, y^0) and (x^1, y^1) with weight λ. For $\lambda = \frac{1}{2}$, we get $(x^{\frac{1}{2}}, y^{\frac{1}{2}}) = (1-\frac{1}{2})(x^0, y^0) + \frac{1}{2}(x^1, y^1) = \frac{1}{2}(x^0 + x^1, y^0 + y^1)$. In Fig. 3.4, we illustrate the position of $(x^{\frac{1}{4}}, y^{\frac{1}{4}})$ and $(x^{\frac{1}{2}}, y^{\frac{1}{2}})$ in an example.

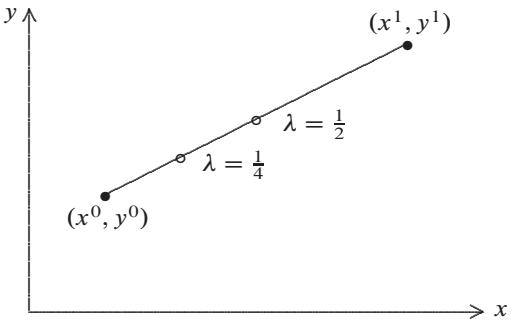


Fig. 3.4 Convex combinations

When we have more firms and thus more observed data points, as shown in Fig. 3.5, we can create not only convex combinations of the original points but also convex combinations of those convex combinations and so on. This gives us the shaded area in the Fig. 3.5.

We can think of any such convex combination of convex combinations as one giant convex combination of more than two points. Let us examine an example with 3 points: (x^1, y^1), (x^2, y^2), and (x^3, y^3), and two convex combinations given by

$$(x^\lambda, y^\lambda) = (1-\lambda)(x^1, y^1) + \lambda(x^2, y^2)$$
$$(x^\mu, y^\mu) = (1-\mu)(x^2, y^2) + \mu(x^3, y^3)$$

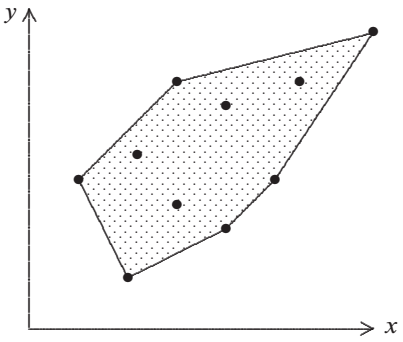


Fig. 3.5 Convex hull

Now consider a convex combination of these combinations

$$(x^\theta, y^\theta) = (1-\theta)(x^\lambda, y^\lambda) + \theta(x^\mu, y^\mu)$$

We will now rewrite this convex combination of the convex combinations. fro simplicity, we focus on the inputs x^θ. The combination can be rewritten as follows:

$$\begin{aligned} x^\theta &= (1-\theta)x^\lambda + \theta x^\mu \\ &= (1-\theta)\big((1-\lambda)x^1 + \lambda x^2\big) + \theta\big((1-\mu)x^2 + \mu x^3\big) \\ &= (1-\theta)(1-\lambda)x^1 + (1-\theta)\lambda(x^2 + \theta(1-\mu)x^2 + \theta\mu x^3 \\ &= (1-\theta)(1-\lambda)x^1 + \big((1-\theta)\lambda + \theta(1-\mu)\big)x^2 + \theta\mu x^3. \end{aligned}$$

Note that the sum of the coefficients of the x'es is

$$\begin{aligned} &(1-\theta)(1-\lambda) + \big((1-\theta)\lambda + \theta(1-\mu)\big) + \theta\mu \\ &\qquad = (1-\theta)(1-\lambda+\lambda) + \theta(1-\mu+\mu) = 1-\theta+\theta = 1. \end{aligned}$$

We can therefore change the notation and write

$$x^\theta = \theta^1 x^1 + \theta^2 x^2 + \theta^3 x^3$$

where

$$\theta^1 + \theta^2 + \theta^3 = 1, \qquad \theta^1 \geq 0, \qquad \theta^2 \geq 0, \qquad \theta^3 \geq 0.$$

and say that x^θ is a convex combination of x^1, x^2, and x^3.

We can therefore write the technology set T based on K observations from firms and the assumption of convexity as

$$\begin{aligned} T = \big\{ (x,y) \,|\, & x = \lambda^1 x^1 + \cdots + \lambda^K x^K \text{ and} \\ & y = \lambda^1 y^1 + \cdots + \lambda^K y^K \text{ for} \\ & \lambda^1 + \cdots + \lambda^K = 1 \text{ and } (\lambda^1, \ldots, \lambda^K) \geq 0 \big\}. \end{aligned} \tag{3.3}$$

or, even more simply, as

$$T = \Big\{ (\sum_{k=1}^{K} x^k, \sum_{k=1}^{K} y^k) \mid \sum_{k=1}^{K} \lambda^k = 1, \lambda^k \geq 0, k = 1, \ldots, K \Big\}. \qquad (3.4)$$

This is the *smallest convex set that contains the K obervations* and is called the *convex hull* of the data set $\{(x^1, y^1), (x^2, y^2), \cdots, (x^K, y^K)\}$.

Convexity is a strong assumption that is often debated in applications and in the theoretical literature. In the DEA literature, for example, several relaxations have been proposed, some of which we will discuss later in the book.

One of the motivations for the convexity assumptions in *microeconomics* is mathematical convenience. Indeed, convexity is required for many of the microeconomic key results that we often rely on. With convex sets, prices are useful controls and offer a dual representation based on separating hyperplanes.

Other more basic motivations include the following:

- Convexity occurs naturally in some contexts. In particular, it occurs when different processes are available and the organization can decide how much time and other resources to allocate to the different processes.
- Convexity provides a reasonable approximation in some contexts. In particular, if the data available on a given firm aggregate data on the processes used in different subunits or subintervals, convex combination can approximate alternative but non-observed aggregations.
- Convexity is sometimes an operationally convenient but harmless assumption as far as results are concerned. This is the case, for example, when we focus on cost efficiency, revenue efficiency and profit efficiency in a setup with fixed prices. In such cases, the results do not change if we invoke the minimal convex set.

From a theoretical and an applied point of view, however, the convexity assumption is not unquestionable. The problems with global convexity assumptions include the following:

- Convexity requires divisibility (because a convex combination is basically an addition of down-scaled plans). This may not be possible when different investments are considered, for example, or when set-up times and switching costs are taken into account.
- Convexity does not take into account the economies of scale and scope (specialization) that are present in many industries.
- Prices may depend on quantity, and thereby, the introduction of convexity is not a harmless convenience.

From a *benchmarking perspective*, convexity allows us to interpolate from observed firms to firms with input-output profiles between the observations. Convexity thereby extends the technology, which in turn enables us to rely on fewer observations and still attain interesting results where not all firms are at the frontier with nothing to learn. On the other hand, it also becomes less obvious which other firms

a given firm can learn from, and we may end relying on a priori assumptions rather than real observations. All of the classical DEA and SFA models presume convexity.

3.6 Free disposal and convex

When we combine the assumptions of free disposability and convexity, we obtain the shaded area in Figure 3.6.

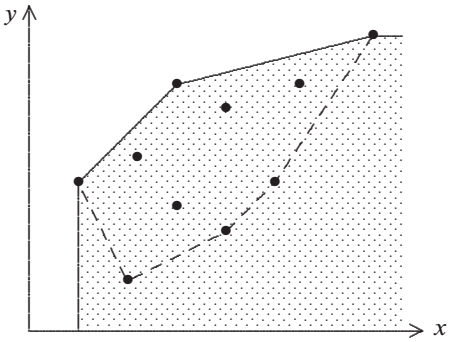


Fig. 3.6 Convex technology set with free disposability

The *convex and free disposal hull* technology derived from K observations can–consistent with the Eq. (3.3)–be written as

$$\begin{aligned} T = \{ (x, y) \mid x \geq \lambda^1 x^1 + \cdots + \lambda^K x^K \text{ and} \\ y \leq \lambda^1 y^1 + \cdots + \lambda^K y^K \text{ for} \\ \lambda^1 + \cdots + \lambda^K = 1 \text{ and } (\lambda^1, \ldots, \lambda^K) \geq 0 \} \end{aligned} \tag{3.5}$$

or, in a more condensed form, as

$$T = \{ (x, y) \mid x \geq \sum_{k=1}^{K} x^k,\ y \leq \sum_{k=1}^{K} y^k,\ \sum_{k=1}^{K} \lambda^k = 1,\ \lambda^k \geq 0,\ k = 1, \ldots, K \}.$$

We see that we form convex combinations as in Eq. (3.3) but that we do not require (x, y) to precisely match this convex combination because we also have disposability, which means that we only need weakly more input x and weakly less output y to ensure feasibility. The set in Eq. (3.5) can be proven to be the smallest set containing data that is both convex and free disposable.

Numerical example

For readers who have never worked with convexity, the above formulations may seem abstract. We therefore provide a small numerical example. Consider the observations in Table 3.4. The observations are plotted in Fig. 3.7. Here we have also

Table 3.4 Numerical input and output for 4 firms

Firm	Input	Output
1	100	75
2	200	100
3	300	300
4	500	400
5	400	200
6	400	375

shown the smallest possible convex and free disposable technology determined from firms 1-4. Firms no. 5 and 6 are extra firms to which we will return. The formula for

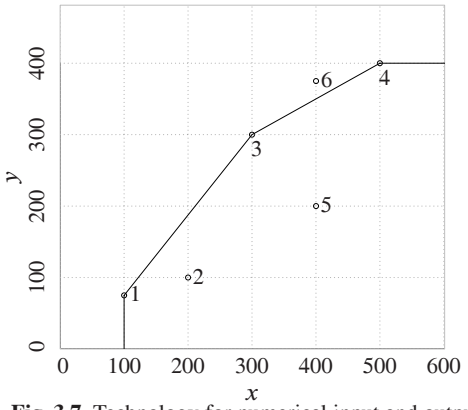


Fig. 3.7 Technology for numerical input and output

the technology formed by Firms 1-4 by assuming convexity and free disposability is as follows (cf. also Eq. (3.5)),

$$\begin{aligned} T = \{\,(x,y) \mid\; & x \geq \lambda^1 100 + \lambda^2 200 + \lambda^3 300 + \lambda^4 500 \\ & y \leq \lambda^1 75 + \lambda^2 100 + \lambda^3 300 + \lambda^4 400 \\ & \lambda^1 + \lambda^2 + \lambda^3 + \lambda^4 = 1, \lambda^1 \geq 0, \ldots, \lambda^4 \geq 0\,\}. \end{aligned}$$

Firm 1 is in this technology set, as can be seen by letting $\lambda^1 = 1$ and all other λs equal zero. Firm 2 is in the technology set based on a similar argument, but Firm 2 is also in the set for other values of λ. For example, $\lambda^1 = 0.6$, $\lambda^3 = 0.4$, and

$\lambda^2 = \lambda^4 = 0$ gives the (in)equalities

$$\begin{aligned} 200 \geq & 0.6 \cdot 100 + 0.4 \cdot 300 = 180 \\ 100 \leq & 0.6 \cdot 75 + 0.4 \cdot 300 = 165 \\ & 0.6 + 0.4 = 1 \end{aligned}$$

that are fulfilled.

Firm 5, with $x = 400$ and $y = 200$, is also in the Technology set; for $\lambda^3 = 1$, we get

$$\begin{aligned} 400 \geq & 1 \cdot 300 \\ 200 \leq & 1 \cdot 300. \end{aligned}$$

On the other hand, Firm 6, with $x = 400$ and $y = 375$, is not in the technology set. To see this, consider the use of convex combinations of Firms 3 and 4. This is our best chance of identifying inequalities that hold. We obtain

$$\begin{aligned} 400 \geq & \lambda^3 \cdot 300 + \lambda^4 \cdot 500 \\ 375 \geq & \lambda^3 \cdot 300 + \lambda^4 \cdot 400 \\ & \lambda^3 + \lambda^4 = 1 \\ & \lambda^3 \geq 0, \lambda^4 \geq 0 \end{aligned}$$

It is clear that there are no values of λ^3 and λ^4 that fulfill these conditions. (For example, one might use the equation to obtain $\lambda^3 = 1 - \lambda^4$ and substitute this into the first inequality. Moving terms around, we find that $\lambda^4 \leq 0.5$. Now, we can substitute $\lambda^3 = 1 - \lambda^4$ into the second inequality; again, a few manipulations give us $\lambda^4 \geq 0.75$. These two requirements are inconsistent, and thus, the inequalities are inconsistent; i.e., observation 6 is not in the technology set, as the graphical illustration also shows.

3.7 Scaling and additivity

A last class of assumptions commonly introduced in both economics and benchmarking concerns the option of scaling operations. It seems likely that if some production plan is feasible, then we can also use somewhat fewer inputs to produce somewhat fewer outputs and slightly increased inputs to produce slightly increased outputs.

More formally, if $(x, y) \in T$, then we should also expect $\lambda(x, y) \in T$ for values of λ close to 1. Now, the question becomes what values of λ we can use.

At one extreme, we have the assumption of *constant returns to scale (crs)* if any possible production combination can arbitrarily be scaled up or down: that is, if $\lambda(x, y) \in T$ for any $(x, y) \in T$ and $\lambda \geq 0$; i.e.,

$$(x, y) \in T,\ \lambda \geq 0 \Rightarrow \lambda(x, y) \in T.$$

Graphically, constant returns to scale mean that when (x, y) is feasible, then any point on a ray from $(0, 0)$ that passes through (x, y) is feasible. When we also assume free disposability, the result is the shaded area in Fig. 3.8.

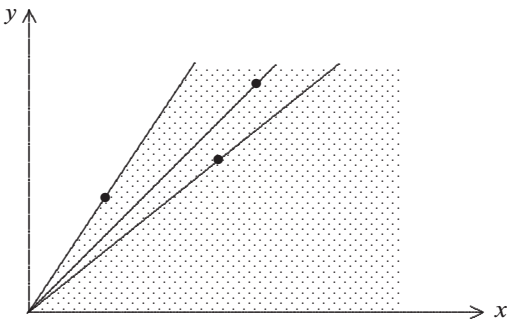


Fig. 3.8 Constant returns to scale

A less extreme assumption is that of *non-increasing returns to scale (nirs)* or (to put it in a way that is slightly less precise but easier to say and to understand) *decreasing returns to scale (drs)*. This situation prevails if for any possible production process, we can arbitrarily decrease the scale of the operation; i.e.,

$$(x, y) \in T,\ 0 \leq \lambda \leq 1 \Rightarrow \lambda(x, y) \in T.$$

Decreasing returns to scale mean that the output will tend to increase less than the input such that it will be possible to scale down but not up. Reasons to expect decreasing returns to scale include whether a firm can run a process at reduced speed, reduce capacity utilization or reduce the amount of time that the process takes.

Graphically, this means that for a given production plan, all plans on the line between zero (i.e., the origo) and this plan are also feasible. This is illustrated in Fig. 3.9, where the technology based on three observations and the assumption of non-increasing returns to scale is composed of the three line segments. If there is also free disposability, the technology set is that indicated in the shaded area.

Instead of assuming that we can scale down but not up, we might assume that we can scale up but not down. This leads to what we naturally might call *non-decreasing returns to scale (ndrs)* or, slightly less precise, *increasing returns to scale (irs)*. This situation prevails if for any possible production process we can arbitrarily increase the scale of the operation; i.e.,

$$(x, y) \in T,\ \lambda \geq 1 \Rightarrow \lambda(x, y) \in T$$

Increasing returns to scale mean that the output will tend to grow faster than the input. One reason for this is that a larger scale implies more experience, more efficient processes and a better ability to utilize specialization possibilities.

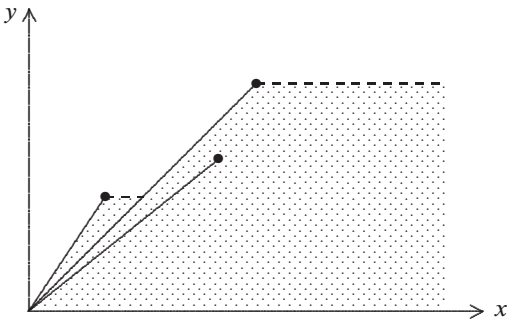


Fig. 3.9 Non-increasing returns to scale

Graphically, this means that for a given production plan, all plans on the line extending from the point but on the same ray compared to zero as the point are also feasible. This is illustrated in Fig. 3.10, where the technology based on the same three observations and the assumption of non-decreasing returns to scale is composed of the three line segments. If there is also free disposability, the technology set is that indicated by the shaded area.

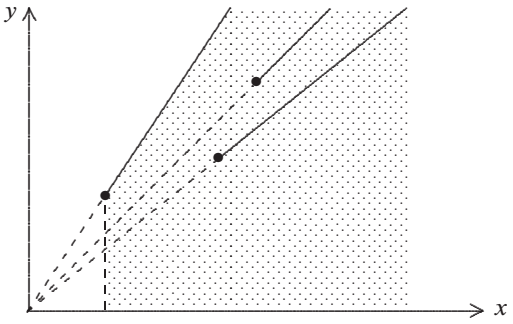


Fig. 3.10 Non-decreasing returns to scale

We close this discussion of rescaling by a slightly different assumption one can naturally make, namely that of *additivity or replicability*. When we have two possible production plans we can look at the sum of the two plans. If we do nothing else it seems plausible that the sum of the two is also possible. This is the assumption of additivity which formally can be expressed as

$$(x, y) \in T,\ (x', y') \in T \Rightarrow (x + x', y + y') \in T$$

The role of additivity is illustrated in Fig. 3.11. Here, to simplify the picture, we have called the two observed input–output combinations a and b rather than the longer (x^1, y^1) and (x^2, y^2).

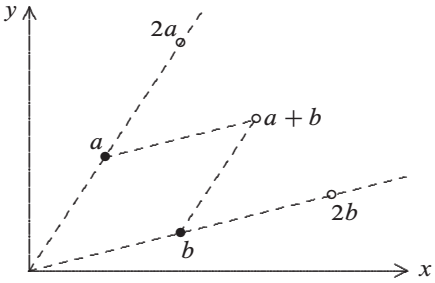


Fig. 3.11 Additivity

Note that additivity also implies that if (x, y) is feasible, so is $2(x, y) = (x, y) + (x, y)$ and therefore also $3(x, y) = 2(x, y) + (x, y)$ and so on. Likewise, if (x, y) and (x', y') are possible, so is $h(x, y) + k(x', y')$ for arbitrary h and k values in $0, 1, 2, 3, \ldots$ (i.e., arbitrary natural number $h \in \mathbb{N}$ and $k \in \mathbb{N}$). We therefore get a full grid of feasible production plans even if we have only observed two such plans to begin with.

Additivity is an appealing assumption because one can think of the added plans as having been executed by running two autonomous production lines or firms next door, one following the first plan and the other the second. The additivity assumption basically rules out positive or negative externalities between the two production plans. Conceptually, therefore, additivity is an appealing assumption. Unfortunately, models based on additivity is somewhat more difficult to implement. In a mathematical programming context, for example, we may have to use mixed integer programming to represent this property.

Of course, there are some relationships between the different regularities that we have introduced. Let us make a few observations that also show the potential power of the additivity assumption. If we assume both non-increasing returns to scale and additivity, then we can just as well assume convexity and constant returns to scale. If we assume convexity and additivity, then we also have constant returns to scale.

Practical application: Waterworks

The interactive benchmarking system IBEN, implemented by the Danish Water and Waste Water Association (see also page 3), allows the use of several technologies, including all of the technologies discussed in this chapter. In addition, a special variant of free disposability is assumed that is called FDH+. One way to look at this assumption is to say that it combines the free disposability assumption with an assumption of *restricted constant return to scale* in the sense that

$$(x, y) \in \{(x^1, y^1), (x^2, y^2), \ldots, (x^K, y^K)\},\ L \leq \lambda \leq H \Rightarrow \lambda(x, y) \in T$$

where $L \leq 1$ and $H \geq 1$ are two numbers that are not very far from 1. Thus, it is assumed that if a production plan is feasible, then we can also use slightly fewer inputs to produce slightly fewer outputs and slightly more inputs to produce slightly more outputs.

The technology set resulting from our three observations in the previous figures, when we assume restricted constant returns to scale and free disposability, is illustrated in Fig. 3.12. Here, we have assumed that L=0.8 and H=1.2, i.e. we assume constant returns to scale as long as we only rescale with 20 % or less.

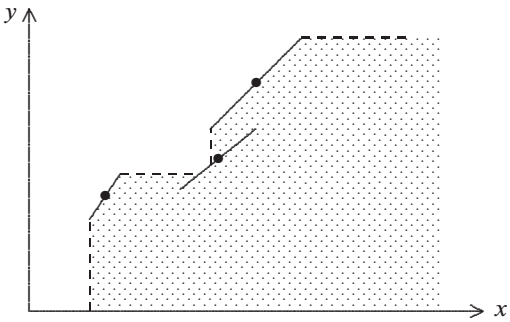


Fig. 3.12 The FDH+ method allowed in Danish Waterworks

Theoretically, this property may not be appealing, and it may even seem internally inconsistent because we do not fully use the rescaling option—we only allow the rescaling of actual observations and not that of observations formed using the free disposability of existing observations. Still, from the point of view of applications, this property has considerable appeal. It allows the user to identify specific existing firms to imitate under the plausible condition that a firm can be resized to a limited degree without really changing the organization and the mode of operation.

3.8 Alternative descriptions of the technology

So far we have described the technology by the set T defined as the set of input–output combinations that we consider to be feasible in the given business case

$$T = \{ (x, y) \in \mathbb{R}^m_+ \times \mathbb{R}^n_+ \mid x \text{ can produce } y \} .$$

We have taken an empirical approach and illustrated how to get an idea of T by combining given observations with simple interpolations and extrapolations from such observations. The principles used are founded on production economics concepts like free disposability, convexity, and return to scale.

In the development of models, computational methods, and applications, it is sometimes useful to use alternative ways to characterize the technology. We will therefore close this chapter with a few observations on such alternatives.

It is sometimes convenient to describe the technology from the input or output side. In such cases, we let the $x \rightarrow P(x)$ and $y \rightarrow L(y)$ be the corresponding *production* and *consumption correspondences*

$$P(x) = \{\, y \mid (x, y) \in T \,\}$$
$$L(y) = \{\, x \mid (x, y) \in T \,\}.$$

The production correspondence is sometimes called the output possibility set or just the output set, and the consumption correspondence is sometimes called the input requirement set or just the input set.

It is clear that if we know $P(\cdot)$ for all values of x, or $L(\cdot)$ for all values of y, we can also (re)construct T. Hence, the input and output correspondences give alternative ways to describe the same technology.

In illustrations, we are often interested in isoquants, i.e. the upper and lower boundaries of $P(x)$ and $L(y)$ respectively. Formally, they are defined as

$$\textit{Isoquant}\, P(x) = \{\, y \in P(x) \mid \theta y \notin P(x) \text{ for } \theta > 1 \,\}$$
$$\textit{Isoquant}\, L(y) = \{\, x \in L(y) \mid \theta x \notin L(y) \text{ for } \theta < 1 \,\}.$$

When we only have one output, $n = 1$, we can define a production function f as

$$y = f(x) = \max\{\, y \mid (x, y) \in T \,\}.$$

If, on the other hand, we have a production function $f(x)$, we can define the technology set T by

$$T = \{\, (x, y) \in \mathbb{R}^m_+ \times \mathbb{R}^n_+ \mid y \leq f(x) \,\}.$$

For a Cobb-Douglas production function for example the technology set is

$$T = \left\{ (x_1, x_2, y) \in \mathbb{R}^3_+ \mid y \leq x_1^{\alpha_1} x_2^{\alpha_2} \right\}.$$

Notice that this technology set fulfills the assumptions of free disposability if f is weakly increasing, i.e. if $x' > x$ implies that $f(x') \geq f(x)$. In a similar way, we can introduce properties of f that correspond to the other general properties of T above. If for example f is concave, we get a convex production possibility set T as illustrated in Fig. 3.13.

Thus it makes no difference whether we describe the technology by a technology set or by a production function. We can use the representation that is most convenient in the specific application.

If there is more then one output, $n > 1$, i.e. in a general multi-input multi-output production structure, the production function approach is less useful since we have to describe the substitution between the different outputs that can typically be produced by a given input vector. In such cases, we will either work directly with the

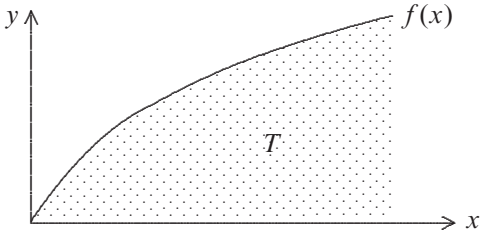


Fig. 3.13 Production function and technology set

technology set or we will use a functional representation based on other function types.

One such possibility is to use distance functions. In Chap. 2 we introduced the Farrell and Shaphard distance functions as measures of efficiency, i.e. we used the distance function to gauge performance. We can, however, also look at distance functions as a way to describe the technology. The idea is simple. Productions plans are feasible when they do not over-perform. Recall from Sect. 2.4 that the Farrell input distance function E and output distance function F are defined as

$$E = E(x, y) = \min\{ E > 0 \mid (Ex, y) \in T \}$$
$$F = F(x, y) = \max\{ F > 0 \mid (x, Fy) \in T \}$$

i.e., E is the maximal proportional contraction of all inputs x that allows us to produce y and F is the maximal proportional expansion of all outputs y that is feasible with the given inputs x. Now, these distance functions provide an alternative description of the technology. Note in particular that if we know $E(x, y)$ or $F(x, y)$ for all $(x, y) \in \mathbb{R}^n_+ \times \mathbb{R}^m_+$, we essentially know T as well. Thus, each of these functions provides a complete characterization of the technology T because

$$E(x, y) \leq 1 \Leftrightarrow (x, y) \in T$$
$$F(x, y) \geq 1 \Leftrightarrow (x, y) \in T$$

or, to put it differently,

$$T = \{ (x, y) \in \mathbb{R}^m_+ \times \mathbb{R}^n_+ \mid E(x, y) \leq 1 \}$$
$$T = \{ (x, y) \in \mathbb{R}^m_+ \times \mathbb{R}^n_+ \mid F(x, y) \geq 1 \}.$$

In the case of the Shephard distance functions

$$D_i(x, y) = \max\{ D > 0 \mid \left(\frac{x}{D}, y\right) \in T \}$$
$$D_o(x, y) = \min\{ D > 0 \mid \left(x, \frac{y}{D}\right) \in T \}.$$

we of course get parallel results

$$D_i(x, y) \geq 1 \Leftrightarrow (x, y) \in T$$
$$D_o(x, y) \leq 1 \Leftrightarrow (x, y) \in T$$

and

$$T = \{ (x, y) \in \mathbb{R}^m_+ \times \mathbb{R}^n_+ \mid D_i(x, y) \geq 1 \}$$
$$T = \{ (x, y) \in \mathbb{R}^m_+ \times \mathbb{R}^n_+ \mid D_o(x, y) \leq 1 \}.$$

We can also represent a general multiple inputs multiple outputs production structure by other functional forms. Most notably, if the technology is convex, we can use what is commonly called *dual representations* of the technology using *cost functions*, *revenue functions* or *profit functions*. The details are given in Appendix 3.11.

3.9 Summary

One of the characteristics of advanced benchmarking studies is that they compare observed performance against a systematic description of possible performances. The latter is provided by what we call the technology set, and a good feel for the nature of technologies is therefore important in benchmarking. The technology set is a description of the input-output combinations that we assume are feasible in a given context. To describe technologies, it is therefore important to understand which assumptions one can reasonably make, explicitly or implicitly, in the construction of technologies based on actual observations.

In this chapter, we have covered the three main classes of assumptions: disposability, convexity and returns to scale. Most importantly, we have defined and illustrated in some detail the following key properties:

Free disposability of input. We can produce the same output with more input, i.e. $(x, y) \in T, x' \geq x \Rightarrow (x', y) \in T$.

Free disposability of output We can produce less output with the same input, i.e. $(x, y) \in T, y' \leq y \Rightarrow (x, y') \in T$.

Free disposability. We can produce less with more, i.e. $(x, y) \in T, x' \geq x, y' \leq y \Rightarrow (x', y') \in T$

Convex. Any weighted average of feasible production is also feasible, i.e. $(x^0, y^0) \in T, (x^1, y^1) \in T, 0 < \lambda < 1 \Rightarrow \left((1-\lambda)x^0 + \lambda x^1, (1-\lambda)y^0 + \lambda y^1\right) \in T$.

Constant returns to scale. Production can be arbitrarily scaled up and down, i.e. $(x, y) \in T, 0 \leq \lambda \Rightarrow \lambda(x, y) \in T$.

Non-increasing returns to scale. "decreasing returns to scale". Production can be scaled arbitrarily down, i.e. $(x, y) \in T, 0 \leq \lambda \leq 1 \Rightarrow \lambda(x, y) \in T$.

Non-decreasing returns to scale. "increasing returns to scale". Production can be scaled arbitrarily up, i.e. $(x, y) \in T, \lambda \geq 1 \Rightarrow \lambda(x, y) \in T$.

Most benchmarking methods presume free disposability and convexity. In addition, some assumption regarding economies of scale is commonly invoked. In applied and less advanced benchmarking studies, it is not always explicitly stated which assumptions are used, but this is important to investigate and understand because it affects the plausibility of the benchmarks we derive.

In addition to the above assumptions, most methods invoke other regularities. Some have economic content, whereas others are invoked for mathematical convenience.

We also discussed alternative equivalent ways to model the technology. Instead of using sets, we can use input and output correspondences. Also, we can use distance functions and when there is only one output, production functions. In the Appendix we will discuss dual representations based on cost functions, revenue functions, and profit functions.

3.10 Bibliographic notes

The notions of disposability, convexity and returns to scale are standard in production theory and benchmarking and are therefore covered in a large number of textbooks. Good modern textbooks include for example Chambers (1988), Färe and Primont (1995), Rasmussen (2010), and Varian (1992).

Additivity and replicability are less common, but have been emphasized by Tulkens (1993), among others.

For more on the bull project see Bogetoft and Nielsen (2004), the credit union project, see Bogetoft et al (2004) and Credit Union National Association (2010), the university project, see Bogetoft et al (2007b), and the waterworks project, see e.g. www.ibensoft.com.

The duality of directional distance functions is proven in Luenberger (1992), which generalizes earlier formulations of Shephard's input duality theorem as shown by Chambers et al (1996)

3.11 Appendix: Distance functions and duality

As mentioned in Sect. 3.8, we can represent a general multiple inputs multiple outputs production structure by alternative functional forms. Most notably, if the technology is convex, we can use what is commonly called *dual representations* of the technology using *cost functions*, *revenue functions* or *profit functions*. A few more details are given in this appendix.

Initially, let us note that distance functions have some useful homogeneity properties. Thus, Farrell input efficiency functions $E(x, y)$ are homogeneous of degree -1 in x, as can be seen in the following computations for $t > 0$:

$$\begin{aligned} E(tx, y) &= \min_E\{\, E \mid (Etx, y) \in T \,\} \\ &= \min_e\{\, \frac{e}{t} \mid (ex, y) \in T \,\} \qquad (e = Et) \\ &= t^{-1} \min_e\{\, e \mid (ex, y) \in T \,\} \\ &= t^{-1} E(x, y)\,, \end{aligned}$$

and the Farrell output distance function is homogeneous of degree -1 in output y,

$$F(x, ty) = t^{-1} F(x, y)$$

as can be shown just like we did above for E.

In the case of Shephard distance functions, we have correspondingly that $D_i(x, y)$ is homogeneous of degree 1 in x and that $D_o(x, y)$ is homogenous of degree 1 in outputs

$$D_i(tx, y) = tD_i(x, y)\,, \quad D_o(x, ty) = tD_o(x, y\,. \qquad (t > 0)$$

The distance functions and the technology set are different ways to describe technological restrictions, and we have seen that they under certain assumptions are equivalent. This might not be a surprise as both methods deal with input and output quantities.

It might come as more of a surprise that the same kind of *duality* exists between the cost, revenue and profit functions and the technology.

The *cost function* is defined as

$$c(w, y) = \min_x \{\, wx \mid (x, y) \in T \,\}$$

where $w \in \mathbb{R}_m$ is a vector of input prices. By use of separating hyperplanes from convex analysis it can be shown that if T is convex then

$$T = \{\, (x, y) \in \mathbb{R}^m_+ \times \mathbb{R}^n_+ \mid wx \geq c(w, y) \text{ for all } w \in \mathbb{R}^m_+ \}\,.$$

By combining with the above we then also have the duality between distance functions and cost functions

$$\begin{aligned} c(w, y) &= \min_x \{\, wx \mid D_i(x, y) \geq 1 \,\} \\ D_i(x, y) &= \min_w \{\, wx \mid c(w, y) \geq 1 \,\}\,. \end{aligned}$$

The same kind of duality also exists for the *revenue function*

$$r(x, p) = \max_y\{\, py \mid (x, y) \in T \,\}$$

and *profit function*

$$\pi(w, p) = \max_{x,y}\{\, py - wx \mid (x, y) \in T \,\}$$

where $p \in \mathbb{R}^n_+$ is a vector of output prices. Thus, if T has free disposability and is convex then

$$T = \{\, (x, y) \in \mathbb{R}^m_+ \times \mathbb{R}^n_+ \mid py \le r(x, p) \text{ for all } p \in \mathbb{R}^n_+ \,\}$$
$$T = \{\, (x, y) \in \mathbb{R}^m_+ \times \mathbb{R}^n_+ \mid py - wx \le \pi(w, p) \text{ for all } (w, p) \in \mathbb{R}^m_+ \times \mathbb{R}^n_+ \,\}.$$

Dual relationships exist also for the directional distance approach and the excess function mentioned in Sect. 2.5. When T is convex and $wd > 0$, we have

$$c(y, w) = \min_x \{\, wx - e(x, y, d)dw \mid (x, y) \in T \,\}$$

and

$$e(x, y, d) = \min_w \{\, wx - c(y, w) \mid wd = 1 \,\}$$

Hence, $e(x, y, d)dw$ may be interpreted as the additional cost of producing y via x rather than in the cost-minimal way.

From a theoretical perspective, the technology set is a purely descriptive concept as are the distance functions and the related efficiencies. It is remarkable that the technology set can be derived from a profit function that normally would be considered a normative concept. The same goes for the efficiency terms; they are descriptive by nature, and it is only by interpretation and introduction of preferences they become normative. Therefore the normative nature of efficiency is only by an implicit introduction of preferences as was discussed in Sect. 1.6.

Chapter 4
Data Envelopment Analysis DEA

4.1 Introduction

In this and the next chapter, we cover the basics and some additional material on Data Envelopment Analysis (DEA). DEA combines the estimation of the technology with the measurement of performance as related to this technology. It thereby integrates the two basic problems of a) defining a performance standard, the technology, and b) evaluating achievements against the established standard. There are several DEA methods that differ in terms of the estimated technology and the efficiency concept used. We will cover the most important ones and emphasize what unites the class of methods that make up DEA.

State-of-the-art benchmarking methods are a combination of two research traditions. One has its origins in management science, mathematical programming and operations research. This is the class of approaches that we refer to as DEA models. The other research tradition has a more economics- and econometrics-oriented background. These are the SFA approaches that we will discuss in later chapters. The two lines of research have lived side by side for many years, each with its group of proponents. The integration of the methods is still limited from a methodological perspective, but researchers from both camps meet regularly at yearly conferences, and more and more researchers and consultants use both types of methods in applications. As we have already indicated in Chap. 1, both approaches have their merits, and it is better to see them as complements rather than as substitutes.

As an OR technique, the DEA approach has gained impressive momentum since it was first proposed in the late seventies. There are now several thousand recorded scientific contributions, some theoretical and some applied.

A short definition of DEA is that it provides a mathematical programming method of estimating best practice production frontiers and evaluating the relative efficiency of different entities. In the DEA literature, these are typically called Decision-Making Units (DMUs), but we will continue to refer to the evaluated entities as firms.

P. Bogetoft and L. Otto, *Benchmarking with DEA, SFA, and R*, International Series in Operations Research & Management Science 157, DOI 10.1007/978-1-4419-7961-2_4,

4.2 Setting

Recall that our general setting involves K firms that use m inputs to produce n outputs. Additionally, let $x^k = (x_1^k, \ldots, x_m^k) \in \mathbb{R}_+^m$ be the inputs used and $y^k = (y_1^k, \ldots, y_n^k) \in \mathbb{R}_+^n$ the outputs produced by firm k, $k = 1, \ldots, K$. We think of these as columns vectors. If input factor prices and output product prices are available, we denote these as $w^k = (w_1^k, \ldots, w_m^k) \in \mathbb{R}_+^m$ and $p^k = (p_1^k, \ldots\ldots, p_n^k) \in \mathbb{R}_+^n$ for firm k. Note that we continue to indicate a firm's identity via a superscript and the individual inputs and outputs via the subscripts.

To condense our notation, we will often write programs in vector form such that a firm's production plan, program, or action is written simply as (x^k, y^k). To simplify the notation, we may drop the superscript when there can be no doubt as to which firm we are considering. If we want to have the production plans for all firms, we will write these in matrix format, $X = (x^1, x^2, \ldots, x^K)$ and $Y = (y^1, y^2, \ldots, y^K)$ Finally, observe that we use $\mathbb{R}_+$ to denote the set of non-negative real numbers. When necessary, we denote the set of strictly positive real numbers as $\mathbb{R}_{++}$.

Lastly, let us introduce the *technology set* or *production possibilities set*

$$T = \{\, (x, y) \in \mathbb{R}_+^m \times \mathbb{R}_+^n \mid x \text{ can produce } y \,\}.$$

The background of the DEA literature is production theory, and the idea is that the firms have a common underlying technology T.

4.3 Minimal extrapolation

Now in reality, we seldom know the technology T. DEA overcomes this problem by estimating the technology T* from observed historical or cross-sectional data on actual production activities. The idea of substituting an underlying but unknown production possibility set with an estimated one is of course not unique to the DEA approach. It is also done in performance evaluations using traditional statistical methods, accounting approaches, etc. What is particular about the DEA approach is the way the approximation of the technology is constructed and the resulting properties of the evaluations. Technically, DEA uses mathematical programming and an activity analyses approach, while the statistical methods are generally based on a maximum likelihood approach. We will return to the technical details and here focus on the main conceptual idea instead.

In DEA, the estimate of the technology T, the empirical reference technology T^*, is constructed according to the *minimal extrapolation principle*: T^* is the smallest subset of $\mathbb{R}_+^m \times \mathbb{R}_+^n$ that contains the data (x^k, y^k), $k = 1, \ldots, K$ and satisfies certain technological assumptions specific to the given approach; for instance, free disposability and some form of convexity. By constructing the smallest set containing the actual observations, the method extrapolates the least.

To formalize the minimal extrapolation principle, we consider candidate technologies T' that are subsets of $\mathbb{R}^m_+ \times \mathbb{R}^n_+$ and that (D) contains data: $(x^k, y^k) \in T'$, $k = 1, \ldots, K$, and (R) satisfy the regularity assumptions. Let the set of such candidate technologies be denoted as

$$\mathcal{T} = \left\{ T' \subset \mathbb{R}^m_+ \times \mathbb{R}^n_+ \mid T' \text{ satisfy (D) and (R)} \right\}.$$

The minimal extrapolation principle means that we estimate the underlying but unknown technology T by the set

$$T^* = \bigcap_{T' \in \mathcal{T}} T'.$$

Under the regularity assumptions, we see that T^* is the smallest set that is consistent with the data. Formally, this follows from the following contradiction argument: if T^* is not the smallest subset, then there exists a $\tilde{T} \subset T^*$ that satisfies (D) and (R). However, then $\tilde{T} \in \mathcal{T}$, and therefore, $T^* = \cap_{T' \in \mathcal{T}} T' \subseteq \tilde{T}$; i.e. we have $\tilde{T} \subset T^*$ and $T^* \subseteq \tilde{T}$—a contradiction.

We can also see that as long as the true technology T satisfies the regularity properties, then $T \in \mathcal{T}$. The approximation that we will develop will be a subset of the true technology, $T^* \subseteq T$. We refer to this as an *inner approximation* of the technology.

It is worth stressing that the minimal extrapolation principle is not applicable with any set of assumptions. It may be that there exist different subsets of $\mathbb{R}^m_+ \times \mathbb{R}^n_+$ containing the observed data and satisfying the assumptions without any possibility of reducing the sets any further. Hence, when developing alternative DEA models, one must prove the existence of the smallest technology, strictly speaking. It is straightforward, however, to do this for any model based on a combination of free disposability, convexity and the standard return to scale properties discussed in Chap. 3. For these properties, it holds that if two sets T^1 and T^2 satisfy all of the conditions, so does $T^1 \cap T^2$. Likewise, if two sets both contain the original observations, so does their intersection. Therefore, a minimal set can be constructed as the intersection of all sets containing data and satisfying the assumptions, and this set will inherit those properties.

Thus, our estimate of the technology set is the smallest possible set that contains the data and fulfills the regularity assumptions. By choosing the smallest set, we are making a *cautious or conservative estimate* of the technology set and therefore also a cautious or conservative estimate of the estimated efficiency and the loss due to inefficiency.

The reliance on inner approximations and the construction of cautious estimates of inefficiency is important in applications. It acknowledges that no observed firm may have reached the frontier of what is technologically feasible, and an approximation based on best practices is therefore cautious. A popular understanding of the property is also that we estimate the technology so as to present the evaluated units in the best possible light—or, as consultants might put it, “we bend ourselves backwards to make you look as good as possible”.

Apart from the sales talk, it is important to understand that the estimated technology set that we use in DEA is founded on the minimal extrapolation principle. It is also important to understand that it is based on the implicit assumption that there is no *noise* in the data. If the data used are somewhat random—due to exogenous chocks, bad reporting practice or ambiguity in accounting practices, for example—the result will not be valid. This may give some substance to a common reaction by firms after an evaluation: namely, that they have been evaluated against the hardest possible standards (possibly the most lucky firms) and not against a cautious standard.

Numerical example

To illustrate why the minimal extrapolation principle only works sometimes, let us consider a simple example in which 4 firms have used one input x to produce one output y as illustrated in Fig. 4.1. Additionally, let us assume that we want to use sets that are delineated by linear (in fact, affine) production functions in the sense that

$$y \leq \alpha + \beta x$$

for some unknown constants α and β. This corresponds to the usual linear models that we construct in the linear regression and the usual accounting view of production and cost functions, where there may be some fixed costs and some variable costs. (Note that the fixed costs here will be $\frac{-\alpha}{\beta}$, while the variable cost per unit of y being produced is $\frac{1}{\beta}$.)

This way of modeling the technology works well with the minimal extrapolation principle. In the figure, we have illustrated three possible sets (below the dashed lines) that have the desired linear regularity and contain the data (the data are below the dashed lines). It is clear that there are many others and that if we take the intersection of all the possible sets, i.e. construct the T^* set, then we get the shaded area. This area, however, is not of the desired nature, i.e. it is not delineated by an affine function.

The problem is that to get a tight fit for the different points, we must select different lines, and there is no single line that simultaneously yields the most conservative evaluation of all of the observations.

Practical application: Regulatory models

From the point of view of incentive provision, as in the regulatory models of DSOs first introduced in Sect. 1.1.3, the minimal extrapolation principle is important as well. Using an inner approximation of the technology and presuming no noise, the regulated firms are able to do at least as well as we stipulate. We might arrive at too high a cost estimate and too low an output estimate, and therefore, the firms might

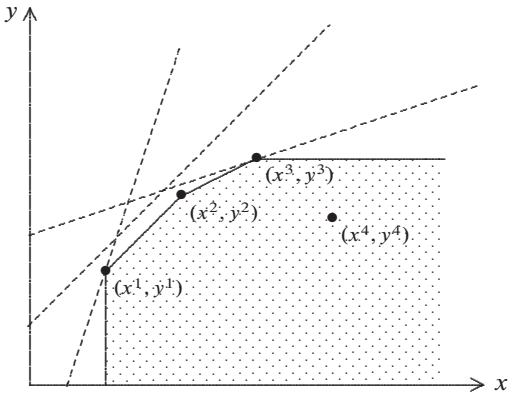


Fig. 4.1 Linear (affine) approximations

earn information rents because they might be able to cut down more on costs or earn higher revenue than we stipulate. This is an important property because we want regulatory systems that are *individually rational*, i.e. that ensure that the firms will participate and not go bankrupt. This explains the tight relationship between minimal extrapolation and individual rationality that we encounter in Chap. 10, where we cover some of the formal results of DEA-based incentive regimes.

4.4 DEA technologies

The basic DEA models mainly differ in the assumptions that they make about the technology T. The most important assumptions are those that that we have already discussed in Chap. 3, where we also provided small illustrations.

We recall the following:

A1 Free disposability. We can produce less with more; that is, $(x, y) \in T, x' \geq x$, and $y' \leq y \Rightarrow (x', y') \in T$

A2 Convexity. Any weighted average of feasible production plans is feasible as well: $(x, y) \in T, (x', y') \in T, \alpha \in [0, 1] \Rightarrow \alpha(x, y) + (1 - \alpha)(x', y') \in T$

A3 γ-returns to scale. Production can be scaled with any of a given set of factors: $(x, y) \in T, \kappa \in \Gamma(\gamma) \Rightarrow \kappa \cdot (x, y) \in T$

A4 Additivity, replicability. The sum of any two feasible production plans is feasible as well: $(x, y) \in T, (x', y') \in T \Rightarrow (x + x', y + y') \in T$

where for γ = crs, drs, irs, or vrs and where the sets of possible scaling factors are given by $\Gamma(\text{crs}) = \mathbb{R}_0$, $\Gamma(\text{drs}) = [0, 1]$, and $\Gamma(\text{irs}) = [1, \infty], \Gamma(\text{vrs}) = \{1\}$, respectively.

The *free disposability* assumption stipulates that we can freely discard unnecessary inputs and unwanted outputs. Except in some cases of joint production (for

instance, where pollution is produced jointly with desirable outputs), this is a safe and weak assumption.

As indicated, in using the term weak, we mean that it is safe to make this assumption because it will most often be fulfilled but also that it contains less power in the sense of extending the production possibility set. Strong assumptions are the opposite.

The *convexity* assumption states that any weighted average (convex combination) of feasible production plans is feasible as well. This assumption is analytically convenient, and some convexity is generally assumed in economic models. Indeed, convexity is necessary for market systems with price-based coordination to work efficiently. Still, convexity is not an innocent assumption, and many attempts have been made in the DEA literature to use weaker-convexity assumptions: e.g., to only assume the convexity of input consumption sets $L(y)$ and output production sets $P(x)$ rather than to assume the convexity of the full set T. In small data sets, convexity has significant power.

The *return to scale* assumptions suggests that some rescaling is possible. Different assumptions have been made regarding the extent and nature of the feasible rescaling. The weakest assumption is that there is no rescaling possible, $\gamma = 1$, and the strongest is that there are constant returns to scale, $\gamma \geq 0$. No rescaling is also called variable returns to scale to produce a common terminology. In between, we may assume that any degree of downscaling is possible but not any degree of upscaling, $\gamma \leq 1$. This means that it cannot be disadvantageous to be small but that it may be disadvantageous to be large, i.e. there may be decreasing returns to scale. More precisely, this concept is sometimes referred to as non-increasing returns to scale. The last and less commonly used assumption, which is actually quite natural and appealing, is that of increasing (or non-decreasing) returns to scale, $\gamma \geq 1$. The idea here is that it cannot be a disadvantage to be large but that it may possibly be a disadvantage to be small.

Lastly, the *additivity* assumption stipulates that when we have feasible production plans, their sum will be feasible as well. Again, this is a natural assumption because one might, for example, imagine that the two original production sites were build next door and run under independent management. By using the original inputs, the sites should therefore be able to produce the same output, and the firm should be able to produce the sum. Unfortunately, additivity is also a difficult assumption to work with and is therefore the least common of the assumptions.

As mentioned above, all DEA models share the idea of estimating the technology using a minimal extrapolation approach, and they only differ in the assumptions that they invoke. In Table 4.1 below, we summarize the assumptions invoked in six classical DEA models: namely, the original constant return to scale (CRS) model; the decreasing, increasing and varying return to scale (DRS, IRS and VRS) models; and the free disposability and free replicability hull (FDH, FRH) models. The latter are not always called DEA models, but we use this terminology because of the common conceptual idea of minimal extrapolation. The last row of Table 4.1 defines some parameter sets Λ that we will use in the construction of the technologies from the actual sets below.

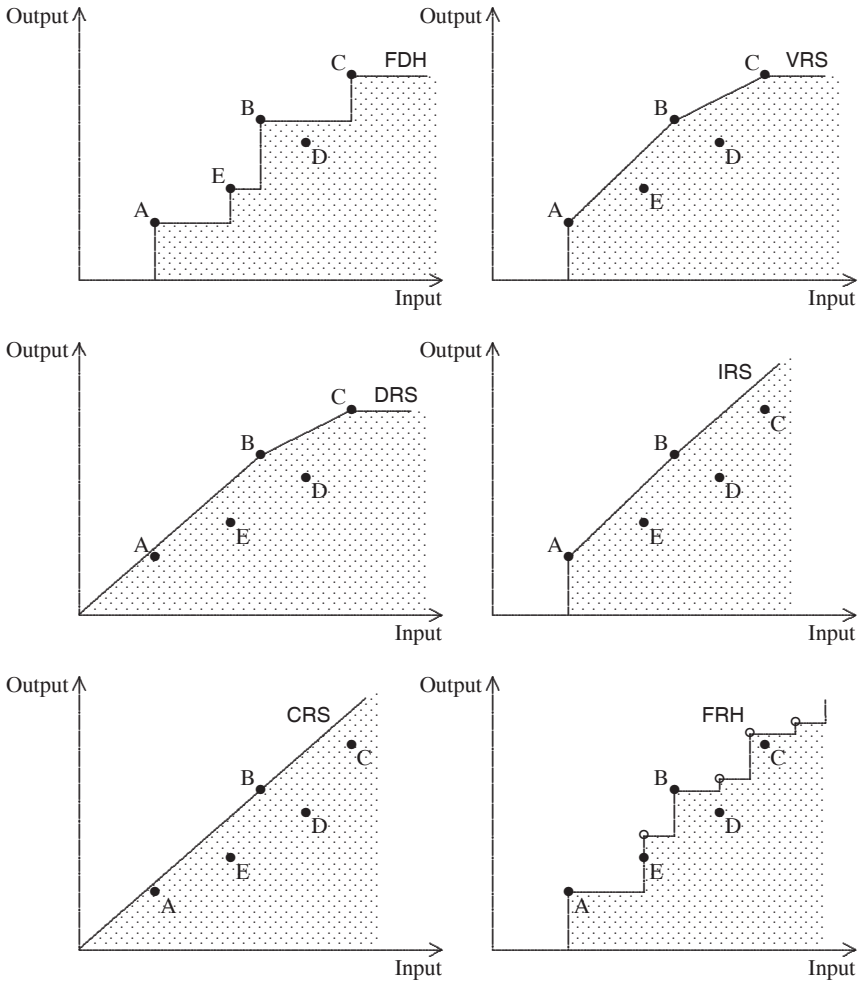


Fig. 4.2 DEA technology sets under different assumptions

In Chap. 3, we have already provided small graphical illustrations of the different assumptions and indicated what we can conclude from the given observations by invoking them individually as well as in combination with one another.

Using the same approach here, we see that the six DEA models in the single–input, single–output case generate technologies like the ones illustrated in Figure 4.2.

We see that the DEA approach involves looking for the smallest set that includes or envelopes the input–output observations for all of the firms. This also explains the name: Data Envelopment Analysis.

Table 4.1 DEA model assumptions

Model	A1 Free disp.	A2 Convexity	A3 γ return	A4 Add.	Parameter set $\lambda \in \mathbb{R}_+^K$
FDH	✓		$\kappa = 1$		$\sum \lambda^k = 1, \lambda^k \in \{0, 1\}$
VRS	✓	✓	$\kappa = 1$		$\sum \lambda^k = 1$
DRS (NIRS)	✓	✓	$\kappa \leq 1$		$\sum \lambda^k \leq 1$
IRS (NDRS)	✓	✓	$\kappa \geq 1$		$\sum \lambda^k \geq 1$
CRS	✓	✓	$\kappa \geq 0$		$\lambda^k \geq 0$
FRH	✓		$\kappa = 1$	✓	$\lambda^k \in \mathbb{N}_0$

It is relatively easy to prove, cf Sect. 4.11, that the minimal extrapolation technologies in the six models are

$$T^*(\gamma) = \left\{ (x, y) \in \mathbb{R}_+^m \times \mathbb{R}_+^n \mid \exists \lambda \in \Lambda^K(\gamma) \colon x \geq \sum_{k=1}^K \lambda^k x^k, \quad y \leq \sum_{k=1}^K \lambda^k y^k \right\}$$

where

$$\Lambda^K(\text{fdh}) = \{\lambda \in \mathbb{N}_+^K \mid \sum_{k=1}^K \lambda^k = 1\}$$

$$\Lambda^k(\text{vrs}) = \{\lambda \in \mathbb{R}_+^K \mid \sum_{k=1}^K \lambda^k = 1\}$$

$$\Lambda^K(\text{drs}) = \{\lambda \in \mathbb{R}_+^K \mid \sum_{k=1}^K \lambda^k \leq 1\}$$

$$\Lambda^K(\text{irs}) = \{\lambda \in \mathbb{R}_+^K \mid \sum_{k=1}^K \lambda^k \geq 1\}$$

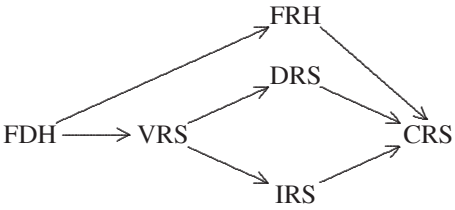
$$\Lambda^K(\text{crs}) = \{\lambda \in \mathbb{R}_+^K \mid \sum_{k=1}^K \lambda^k \text{ free}\} = \mathbb{R}_+^K$$

$$\Lambda^K(\text{frh}) = \{\lambda \in \mathbb{N}_+^K \mid \sum_{k=1}^K \lambda^k \text{ free}\} = \mathbb{N}_+^K$$

and where we have used $\mathbb{N}_+$ to denote the non-negative integers (natural numbers).

It is important to understand that the estimates of the technology, the $T^*(\gamma)$ sets, are derived from the feasibility of the observations and the regularity assumptions using the minimal extrapolation principle. That is, the mathematical set $T^*(\gamma)$ is the smallest set containing data and fulfilling the assumptions that we have listed in the model we call γ.

The way that we express the FDH model may look somewhat complex next to the simplicity of the FDH technology. This complexity serves, however, to allow us to express all of the models in a similar way and thereby to emphasize their relationships. Thus, it is clear from the above formulations that the larger the $\Lambda^K(\gamma)$ set, the larger the estimated technology. It follows that we can partially rank the technologies from smaller to larger (indicated by arrows below) in the following way:



Thus, FDH is the smallest technology set. VRS is larger because we have "filled out the holes". By allowing some scaling, we arrive at a larger set, either DRS (which enlarges the set for small input values) or IRS, which enlarges the technology for the large input values. Allowing full rescaling and convexity, we determine our largest technology, the CRS technology. The FRH is somewhat less comparable to the others, but it is larger than FDH and smaller than CRS. Of course, one can also see this from Table 4.1.

These relationships are interesting because they suggest systematic differences between the outcomes of benchmarking exercises depending on the assumptions that we make a priori. The larger the Γ^K sets, the larger the estimated technology, i.e. the more optimistic we are in estimating the improvement potential of a firm. The flip-side of this is of course that the firms look less efficient in the larger models. Ideally, then, the choice of assumptions shall be carefully argued and, if possible, tested. We will show some such tests in Chap. 6.

Practical application: DSO regulation

In regulatory applications, it is always important to discuss which assumptions to make a priori, and this discussion attracts a great deal of attention. This is not surprising because it may have a huge impact on the revenues that companies are allowed to charge.

Firms therefore normally unanimously prefer the FDH model. In regulations, however, there are seldom enough data to avoid the convexity assumption. Of the VRS, DRS, IRS and CRS models, the firms also tend to prefer the VRS models because they all have higher efficiency scores in this and thereby, for example, higher cost norms. The choice between DRS and IRS is seen as favoring either the large or the small, while the CRS is the worst alternative for most firms.

The regulator, however, should have different interests. To limit the firms' informational rents and as a (partial) representative of the consumers, the regulator will tend to prefer large models to reduce the reasonable charges. However, this is

not the regulator's only concern. The regulator will also not wish to be too harsh because this may reduce relevant maintenance, halt relevant new investments or, in the worst case scenario, drive sound firms into bankruptcy. This may lead the regulator to favor a smaller model. A third concern for the regulator can be structural development. Choosing a CRS model, for example, will give the firms incentives to reorganize, merging ones that are too small and splitting ones that are too large to adjust them to the optimal scale, as we will see in Sect. 4.8 below.

4.5 DEA programs

When we combine the idea of minimal extrapolation with Farrells idea of measuring efficiency as a proportional improvement, we obtain the mathematical programs that many consider synonym with the DEA approach.

On the input side, we measure the Farrell efficiency of firm o as the *input efficiency*

$$E^o = E((x^o, y^o); T^*) = \min\{ E \in \mathbb{R}_+ \mid (Ex^o, y^o) \in T^* \}$$

If we insert the formulation of $T^*(\gamma)$ from above, we get

$$\begin{aligned} \min_{E,\lambda^1,\ldots,\lambda^K} \; & E \\ \text{s.t.} \quad & Ex^o \geq \sum_{k=1}^{K} \lambda^k x^k, \\ & y^o \leq \sum_{k=1}^{K} \lambda^k y^k, \\ & \lambda \in \Lambda^K(\gamma). \end{aligned}$$

For the sake of clarity, let us also write out this somewhat compact vector form in coordinate form as

$$\begin{aligned} \min_{E,\lambda^1,\ldots,\lambda^K} \; & E \\ \text{s.t.} \quad & Ex_i^o \geq \sum_{k=1}^{K} \lambda^k x_i^k, \quad i = 1, \ldots, m \\ & y^o \leq \sum_{k=1}^{K} \lambda^k y_j^k, \quad j = 1, \ldots, n \\ & \lambda \in \Lambda^K(\gamma). \end{aligned} \tag{4.1}$$

Hence, the DEA approach to efficiency measurement leads to a mathematical optimization problem. This explains why DEA is sometimes referred to as the mathematical programming approach to efficiency analyses.

On the output side, we similarly measure the efficiency of firm o as the *output efficiency* using

$$F^o = F((x^o, y^o); T^*) = \max\{ F \in \mathbb{R}_+ \mid (x^o, Fy^o) \in T^* \}$$

and inserting the formulation of T^*, we get the following linear programming problem:

$$\begin{aligned} \max_{F,\lambda^1,\ldots,\lambda^K} \quad & F \\ \text{s.t.} \quad & x^o \geq \sum_{j=1}^{K} \lambda^k x^k \\ & Fy^o \leq \sum_{j=1}^{K} \lambda^k y^k \\ & \lambda \in \Lambda^k(\gamma). \end{aligned}$$

In the case of constant returns to scale, we must confront an inverse relationship between input and output efficiency: i.e. $F = {}^1\!/_{\mathrm{E}}$.

In all of the classical cases we have considered, the optimization problems are relatively simple. They involve $K+1$ variables, a linear objective function, m linear input constraints and n linear output constraints plus possibly an additional linear constraint and possible integer constraints from the $\lambda \in \Lambda^K(\gamma)$ constraint. In the CRS, VRS and DRS cases, the programs are simple *linear programming (LP)* problems, and in the FDH and FRH cases, they are *mixed integer programming (MIP)* problems with integer λ variables. We provide a brief introduction to some key linear programming results in the Appendix 4.11

Although we have formulated the FDH model similar to the other models above, we should note that the FDH model will typically not be solved using MIP routines. In fact, this would be overkill because it is possible to rewrite the program as a series of simple minimax problems that can be solved using a well defined series of simple comparisons. It is easy to see, for example, that the input and output efficiency of (x^o, y^o) relative to the FDH technology is

$$\begin{aligned} E^o(\text{fdh}) &= \min_{k:y^k \geq y^o} \max_{i=1,..,m} \frac{x_i^k}{x_i^o} \\ F^o(\text{fdh}) &= \max_{k:x^k \leq x^o} \min_{j=1,..,n} \frac{y_j^k}{y_j^o} \end{aligned}$$

To understand these formulations, note that to find the input efficiency of a FDH technology, we must look at all of the firms that are producing more of the outputs

to find a relevant comparator. We are looking to find the comparator that makes firm o look the least efficient, so we first minimize the outer optimization. Now, for a candidate peer unit, we must determine which inputs lead to the highest performance evaluation of firm o because this determine the largest proportional reduction that we can make to all inputs at the same time. Then, in the second, inner optimization, we maximize. The logic of the output-based measure in the FDH case is similar. As a result of these formulations, one can easily write routines with two nested loops that find FDH efficiency scores.

Practical application: DSO league tables

As mentioned in Sect. 1.1.3, many European countries routinely benchmark their electricity distribution system operators or DSOs. In Sweden, for example, this started in 2000, where a series of models were analyzed with the aim of establishing useful and reliable efficiency measurements for the more than 200 Swedish DSOs. The results are published on a yearly basis and read with great interest by the companies as well as local politicians and consumers.

An example of what the published results might look like is shown in Fig. 4.3 below. Such tables are the usual primary outputs of benchmarking exercises. In fact, in Sweden, the regulator goes a step further by both offering the results of several models and providing color coding for easy interpretation. The color coding use green as an indication of fine performance, yellow as an indication of performance that should be closer monitored, and red as an indication that performance is clearly unsatisfactory.

In the Fig. 4.3, we see also four columns of efficiency scores. The first three are traditional Farrell input efficiency scores, while the last is a scale efficiency measure that we will return to below. The three Farrell efficiency scores are derived from three different models. Each model describes the production in a DSO as a transformation of different costs types to 5 outputs: Delivered energy High Voltage (MWh), Delivered energy Low Voltage (MWh), No of connections High Voltage, No of connections High Voltage, and Maximum demand (MW). This is done while taking into account 3 environmental conditions, including network length and climate.

The three models differ on the input side, where the short run model (SR) focus on the reduction of the operating expenditures Opex while the long-run model (LR) also focus on reducing Net-losses and capital expenditures, Capex. Lastly, the price-efficiency model (PE) focus on the cost to consumers, which may differ from the firms' cost if the firms have different profit margins. The models also deviate in terms of the assumed return to scale. In the short-run model, it is assumed that a VRS specification is most relevant because the DSOs have limited opportunities to reorganize in the short run. The LR model, the other hand, assume CRS because different concession areas could possibly be integrated over a longer time span.

Company	SR	PE	LR	SE
Hedesunda Elektriska AB	1.00	1.00	1.00	1.00
Nors och Segerstads El.Andelsförening	0.32	0.86	0.71	0.98
Åkab Nät och Skog AB	0.83	0.88	1.00	1.00
Karlsborgs Energi AB	0.67	1.00	0.81	0.95
Edsbyns Elnät AB	0.66	0.86	0.42	0.42
Birka Nät AB,Munkedal	0.67	0.64	0.68	0.95
Herrljunga Elektriska AB	0.70	0.79	0.75	0.86
Björnekulla Energi AB	0.73	0.54	0.73	1.00
Tibro Elverk	0.95	1.00	1.00	1.00
Kramfors Energiverk AB	0.58	0.75	1.00	1.00
Ljusdal Elnät AB	0.84	0.91	0.79	0.79
Kristinehamns Energi Elnät AB	0.59	0.68	0.59	0.99
Birka Nät AB,Hunnebo	0.67	0.70	0.67	0.98
Birka Nät AB, Hudiksvall	1.00	1.00	0.98	0.98
Tranås Energi AB	0.56	0.79	0.59	1.00
Vattenfall Norrnät AB, Kalix	0.86	0.93	0.75	0.84
ARVIKA ELNÄT AB	0.61	0.61	0.90	0.90
Ale Elförening ek för	0.43	0.70	0.54	0.91
KREAB Energi AB	0.62	0.69	0.62	0.85
Lerum Energi AB	0.72	0.87	0.84	0.94
Malungs Elnät AB	0.62	0.73	0.63	0.86
Härjeåns Nät AB	0.71	0.82	0.59	0.59
Vattenfall Sveanät AB, Östra Roslags	0.53	0.84	0.79	0.79
Skövde Elnät	1.00	0.88	1.00	1.00
Bodens Energi Nät AB	0.84	1.00	0.83	0.83
Karlskoga Elnät AB	0.49	0.66	0.51	0.87
Birka Nät AB,Orust/Tjörn	0.60	0.55	0.51	0.85
Leksand-Rättvik Elnät AB	0.58	0.77	0.54	0.79
Ringsjö Energi AB	0.81	0.70	0.72	0.85
AB PiteEnergi	1.00	1.00	1.00	1.00
Birka Nät AB, Strömstad	0.71	0.59	0.56	0.79
Vattenfall Västnät AB, Skaraborg	0.49	0.59	0.41	0.60
Växjö Energi Elnät AB	0.44	0.71	0.52	0.93
AB Borlänge Energi	1.00	0.89	1.00	1.00
Karlstads Elnät AB	0.80	0.78	0.79	0.96

Fig. 4.3 Part of Swedish DSO performance table 2002 (test version)

4.6 Peer units

The right-hand sides in the DEA program Eq. (4.1) defines the *reference unit*

$$\Big(\sum_{k=1}^{K}\lambda^k x^k, \sum_{k=1}^{K}\lambda^k y^k\Big)$$

against which we compare firm o. We see that the DEA program identifies a specific reference unit, most often a weighted average of the existing units and that the reference unit may vary with the evaluated unit. The units with positive weights are typically called *peer units*, i.e.

$$\text{Peer Units} = \big\{k \in \{1,\ldots,K\} \mid \lambda^k > 0\big\}$$

and we can therefore say that *DEA identifies explicit real peer-units for every evaluated unit.*

Graphically, the reference unit is the unit on the technological frontier that firm o is projected onto, and the peer units are the actual frontier units that spans the part of the frontier where the reference unit is located. A numerical example is provided in Sect. 4.6.1.

The reference unit and the associated peer units are usually interpreted as the ones demonstrating how firm o can improve.

Of course, this argument is most convincing when there is actually only one peer unit because it is not clear exactly how to imitate a convex combination, especially when the peers involved are very different in terms of the resource combinations that they use and the service combinations that they deliver. This makes the FDH approach particularly appealing. Additionally, the FRH approach can be thought of in this way because the reference unit in this case can be understood as the sum of existing firms, and this may guide strategic decisions intended to improve firm o.

Although the classical DEA models will typically produce combined reference units, i.e. use weighted averages of figures for several firms, it is still the case that the DEA models will use references based on a much reduced set of firms compared to, for example, the parametric models. One can therefore argue that a distinct advantage of DEA is that it *provides explicit, real peer-units*.

In the DEA models, the *number of possible peer units* for a given firm is equal to the number of inputs plus the number of outputs except in the CRS case, where there can generally be one less peer unit. This follows from the fundamental linear programming results covered in the Appendix 4.11. According to LP theory, if there exists an optimal solution, there exists a basis optimal solution for which the number of positive variables is at most equal to the number of linear restrictions. In the VRS, DRS and IRS programs, there are $m + n + 1$ rows and $K + 1$ variables, (λ and E), and because E is typically positive, there are $m + n + 1 - 1 = m + n$ variables left that may be positive. In the CRS model, we have one constraint less. Although there may be $m + n$ or $m + n - 1$ peer units for a given firm, there will typically be less. This happens when there is slack in the solution. We will return to this below.

A result of the above is that the more inputs and outputs are included in an analysis, the more firms are in the reference set and the more firms have an efficiency of 1. Therefore, one ought only to include inputs and outputs that are definitely relevant. Including too many inputs and outputs will tend to make many firms efficient, i.e. the methods lose their discriminatory power or their ability to distinguish the high performers from the rest. To put it differently, with few data points, we are unable to estimate complex technologies of high dimensionality.

For these reasons, DEA researchers have suggested *rules of thumb* for the relationship between the number of firms and the number of inputs and outputs. The traditional rules are that we need $K > 3(m + n)$ and $K > m \cdot n$, i.e. the number of firms must exceed 3 times the number of inputs plus the number of outputs, and the number of firms must exceed the product of the number of inputs and the number of outputs. These requirements are definitely at the low end, and one can propose other rules: e.g., by comparing to the number of unknown parameters in the most flexible parametric model, the translog model, which we will discuss in Chap. 8.

Practical application: Waterworks

This also explains why peer information is made explicit in several studies. An example is IBEN, the interactive benchmarking approach used by the Danish waterworks, cf. section refsec:IBEN. In IBEN, the peer units and their relative weights,

the λ values, are illustrated by the bars in the lower part of the screen, and users can learn by accessing all available information about the peer units via a separate tab. In fact, on the main screen shown in Sect. 2.5, users can click on the bars to remove peers units that they find less interesting and perform a new analysis of the reduced set of observations. This case emphasizes the importance assigned to the peer units in real applications.

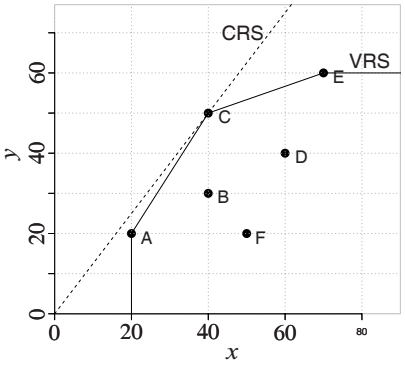
More generally, the peer units for a given firm are often considered a signal of the modeling quality. In real-life instances, we have therefore often had industrial representatives on study steering committees examine a series of peers for different firms to evaluate their relevance. Our practical experience is that if one uses reasonable inputs and outputs, the peer units will most often seem natural for the industrial partners as well.

4.6.1 Numerical example in R

Production data and the corresponding graphs for a case in which six firms have used 1 input to produce 1 output are provided in Table 4.2.

Table 4.2 Data for six firms with one input and one output

firm	input x	output y
A	20	20
B	40	30
C	40	50
D	60	40
E	70	60
F	50	20



Generally, we must formulate the mathematical program and then find a solver to actually do the calculations. Let us illustrate how the general formulations above look in the specific example. To find the input efficiency of firm B using the VRS technology, we must solve the following program, cf. the general formulation in Eq. 4.1,

$$
\begin{aligned}
\min \quad & E \\
\text{s.t.} \quad & E \cdot 40 \geq \lambda^A 20 + \lambda^B 40 + \lambda^C 40 + \lambda^D 60 + \lambda^E 70 + \lambda^F 50, \\
& 30 \leq \lambda^A 20 + \lambda^B 30 + \lambda^C 50 + \lambda^D 40 + \lambda^E 60 + \lambda^F 20, \\
& = \lambda^A + \lambda^B + \lambda^C + \lambda^D + \lambda^E + \lambda^F, \\
& \lambda^A \geq 0, \ldots, \lambda^F \geq 0, E \geq 0.
\end{aligned}
$$

This example is sufficiently simple that we can easily determine the production possibility set and calculate the efficiency figures by hand. However, with just a few more observations and more inputs and outputs, calculations by hand become tedious, and we must rely on computers.

DEA problems can be solved using any of the large number of general linear programming routines available. Standard spreadsheets can even be used to solve a fair number of simple problems. However, such an approach would of course become tedious if not impossible when one had to solve several large problems. It is therefore more convenient to use some of the specialized *software* programs that have already been developed to solve DEA problems. The advantage in general is that they are easy to begin using. The potential drawback is that one is limited to whatever functions the developer has included.

The use of a free open-source software like that developed through the R project can help overcome these limitations. In addition, it allows for easy integration into other mathematical and statistical models. The downside is that coding takes a little longer to learn. Still, with the packages that are available now, it is relatively easy to begin. In this book, we therefore rely on R to illustrate both the DEA and the SFA models.

Moreover, we have developed a package called Benchmarking that contains all of the main DEA and SFA methods that we cover in this book. We must also note that there are other DEA and SFA routines developed for R and that are available through the R repository. Some of them are mentioned and compared in Chap. A, where we also give a short introduction to R and the relevant benchmark packages. We note in particular that Poul Wilson has developed an interesting package called FEAR, Wilson (2008), which is available through his personal homepage. Paul Wilson was a forerunner in the use of R for DEA problems, and his routines have the advantage of conducting massive computations very quickly.

To estimate the input efficiency of the data in the above table, assuming variable returns to scale, VRS, we can use our "Benchmarking" package in R with the following commands:

```
> library(Benchmarking) # load the Benchmarking library
> x <- matrix(c(20, 40, 40, 60, 70, 50),ncol=1) #define inputs
> y <- matrix(c(20, 30, 50, 40, 60, 20),ncol=1) #define outputs
> e_vrs <- dea(x,y, RTS="vrs", ORIENTATION="in")#solve LP problem
> eff(e_vrs) #select efficiency scores from the results in e
```

Output from these commands are

```
[1] 1.0000 0.6667 1.0000 0.5556 1.0000 0.4000
```

showing that most of the firms are fully efficient, as we can also see from the figure in Table 4.2.

Now, via obvious modifications, we can also make calculations under different assumptions. Thus, for example, if we want to determine the Farrell efficiency of the CRS model, we use

```
> e_crs <- dea(x,y, RTS="crs", ORIENTATION="in")
> eff(e_crs)
[1] 0.8000 0.6000 1.0000 0.5333 0.6857 0.3200
```

We see that in the CRS model, only firm C is fully efficient, as the figure also shows.

We can similarly and easily calculate efficiency under alternative assumptions about the technology as well, using the RTS option with values ”drs”, ”irs”, ”fdh” and ”add”, where the latter is what we have also called FRH. The results are summarized in Table 4.3. We see how efficiency falls (or stays constant) when we move towards a larger technology. Thus, FDH efficiency is always the largest and CRS efficiency the smallest for a given firm. Additionally, VRS efficiency is always weakly larger than DRS and IRS efficiency.

Table 4.3 Efficiency for six firms

Firm	Input	Output	E^{FDH}	E^{VRS}	E^{DRS}	E^{IRS}	E^{FRH}	E^{CRS}
A	20	20	1.00	1.00	0.80	1.00	1.00	0.80
B	40	30	1.00	0.67	0.60	0.67	1.00	0.60
C	40	50	1.00	1.00	1.00	1.00	1.00	1.00
D	60	40	0.67	0.56	0.53	0.56	0.67	0.53
E	70	60	1.00	1.00	1.00	0.69	0.86	0.69
F	50	20	0.40	0.40	0.32	0.40	0.40	0.32

It is also very easy to determine the peers and weight information using the `peers` and `lambda` functions in the efficiency calculations.

```
> e_vrs <- dea(x,y, RTS="vrs", ORIENTATION="in", NAMES=TRUE)
> peers(e_vrs)
     [,1] [,2]
[1,]    1   NA
[2,]    1    3
[3,]    3   NA
[4,]    1    3
[5,]    5   NA
[6,]    1   NA
> lambda(e_vrs)
             L1        L3 L5
[1,] 1.0000 0.0000  0
[2,] 0.6667 0.3333  0
[3,] 0.0000 1.0000  0
[4,] 0.3333 0.6667  0
[5,] 0.0000 0.0000  1
[6,] 1.0000 0.0000  0
```

The `peers` function tells us that unit 1 (=A) is being compared with unit 1 (=A), i.e. to itself. This is not surprising because unit 1 is efficient. Firm 2 (=B), however, has two peers, 1(=A) and 3 (=C). The case is similar for the other firms. Note that there are at the most two peers. This is in accordance with the theory we covered above because we have 1 input and 1 output and use a VRS model, i.e. there can be at most 1+1 peer units. The relative importance of the peer units—i.e. the λ values—is extracted with the `lambda` function. We see that there are 3 active peers in total. The λ values that are sometime strictly positive are $L1(= \lambda^A)$, $L3(= \lambda^C)$, and $L5(= \lambda^E)$. We see also that, for example, Firm 4 (=D) is compared to a weighted average of 1 and 3, with 3 accounting for 2/3 of the weight. This result is also clearly depicted in Table 4.2 because firm D is indeed projected on the line segment between A and C and closest to C.

To use the real names of firms instead of numbers, one can make use of the names option in Benchmarking. Often, however, particularly with large data sets and long names, it is convenient simply to number the units and then substitute in the names in the final presentation.

4.7 DEA as activity analysis

Some authors like to conceptualize the DEA model as an activity analysis model with reference to Koopmanns, the first Nobel Prize Winner in Economics (1975). For people trained in linear programming, this makes perfect sense because activity analysis is a very powerful modeling approach that has been used since the 1950s to model real problems using LP problems.

In an activity analysis model, we basically start out by describing the different activities in an organization: e.g., the different machines or processes. These processes are represented by column vectors defining how inputs are transformed into outputs. In a farm model, for example, each cow could be an activity transforming different types of input: foodstuffs, labor and capital into different types of output: milk, calves, manure, etc. Additionally, we could include activities representing different crops. The question asked in activity analysis is how intensely to use the different activities: e.g., how to divide the foods among the cows and how to divide the labor between animals and crops. The constraints in this case will therefore reflect the available resources: e.g., the amount of food available and the balance of the different resources.

It is clear that the DEA problems are similar to such classical operation research models. We just use realized input-output combinations as different columns in the LP problem, and the question of activity intensity becomes one of finding the λ weights. Hence, DEA models are essentially activity analysis models with the added feature that information about the activities is provided via actual observations rather than, for example, expert descriptions of what might be done.

This also points to another novel feature of DEA. In DEA, we use LP to evaluate the past, while traditional OR uses LP to plan the future.

Practical application: Quasi-activities in regulation

This analogy can also guide the combination of DEA with other techniques. If we can make engineers or organizational specialists discover new ways to transform resources into services, we can in principle include them as columns in the DEA problem just in the same way that we do the input–output combinations that we develop. We can than benchmark not only against best practices used but also against possible improvements on best practices.

One area in which this has been done is regulation of network companies. In several countries, regulators are experimenting with the use of engineering models to supplement models derived purely from real observations. If we can predict what a redesigned network using modern equipment may be able to accomplish and then include some such quasi-networks as artificial observation, we may obtain a more forward-looking benchmark.

So far, such efforts have only been experimental, and the main use of the engineering models has been to identify potentially important inputs and outputs that can guide empirical modeling. For example, this occurs in the German regulation of transmission companies as a way to compensate for a small sample. In Chile, regulation also involves some non-realized quasi-observations; there, however, they are developed by different teams of management consultants that investigate different subsets of the firms for possible improvements.

4.8 Scale and allocative efficiency

4.8.1 Scale efficiency in DEA

In the CRS model, and to some extent the DRS and IRS models, the return to scale properties are fixed by assumption. This is not the case for the VRS model, and one may therefore wish to know what will happen if we slightly rescale a firm. One possibility is that the inputs and outputs will be scaled up and down with the same proportions. This corresponds to local constant return to scale. Another possibility is that we can scale the firm up at least slightly but not down based on local decreasing returns to scale. The last possibility is that we can scale up slightly but not down, i.e. there may be local increasing returns to scale.

In a single–input, single–output model VRS model, it is easy to see that as we move along the frontier from smaller to larger inputs, the returns to scale is first increasing, then constant and finally decreasing. Geometrically, this means that a line from (0,0) to a frontier point has a slope that first increases, then stalls, and finally decreases. Economically, it means that the average product—i.e. the number of outputs per input unit–first increases, then is constant and then falls. We call the input level at which we have constant return to scale the *most productive scale size (MPSS)*. At the most productive scale size, the average output is maximal, and in a

single-input cost model, the average costs in minimal. If possible, all firms would like to operate here.

In a multiple–input, multiple–output setting, we see a similar pattern as we traverse the efficient frontier in a given direction in the input and output space, i.e. when we look at the point $(tx, F(tx, y)y)$ as t increases.

Now, to measure the loss from not operating at optimal scale size, we use the notion of *scale efficiency SE*. We calculate this as the ratio of input efficiency in a CRS model to that in a VRS model, i.e.

$$SE(x^o, y^o) = \frac{E(x^o, y^o; crs)}{E(x^o, y^o; vrs)}$$

We see that this measure is never higher than 1 and that it is precisely 1 when the VRS and CRS technologies coincide, i.e. when a firm is operating at optimal scale size. The smaller the value of SE, the more is lost from not having the high average product that one would have at the most productive scale size.

To better understand SE, we can rewrite the above definition as

$$E(x^o, y^o; crs) = E(x^o, y^o; vrs) \cdot SE(x^o, y^o)$$

This means that we can decompose the efficiency (related to a CRS technology) into two components: 1) pure (technical) efficiency measuring the ability to use best practices in the VRS technology and 2) scale efficiency measuring the ability to operate where the average output bundle per input bundle is maximal. A graphical illustration is provided in Fig. 4.4 below. We see that the size of SE can be calculated by comparing the necessary inputs on the efficient VRS frontier and the necessary inputs on the CRS frontier.

$$E(x^o, y^o; crs) = \frac{\|x^{crs}\|}{\|x^o\|} = \frac{\|x^{crs}\|}{x^{vrs}\|} \cdot \frac{\|x^{vrs}\|}{\|x^o\|} = SE(x^o, y^o) \cdot E(x^o, y^o; vrs)$$

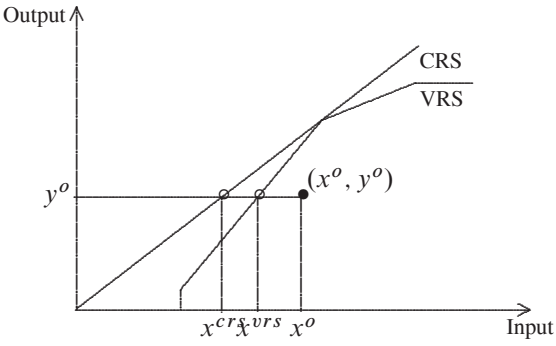


Fig. 4.4 Scale efficiency in VRS technology

Scale efficiency expresses how close the firm is to the optimal scale size: the larger the SE, the closer the firm is to optimal scale. This is interesting information because it indicates the likely gains from adjusting firm scale. Unfortunately, it does not show to what extent an SE less than 1 is due to the firm's being too small or too large. This is easy to determine, however, by also calculating the efficiency under DRS. Thus, if $E(x^o, y^o; drs) = E(x^o, y^o; crs)$, the firm is *below optimal scale size* because here the DRS and CRS technologies coincide. Alternatively, if $E(x^o, y^o; drs) = E(x^o, y^o; vrs)$, the firm is *above optimal scale size*. One can of course arrange this test in different ways. One could for example just look at $E(x^o, y^o; vrs) - E(x^o, y^o; drs)$. If this is 0, the firm is on or above optimal scale size and if it is equal to 0 the firm is on or below optimal scale size. An alternative approach is to look at $\sum_k \lambda^k$. If this sum is less than 1, the firm is below optimal scale size, and if it is above 1, the firm is above optimal scale size.

The idea of looking at scale efficiency is appealing because it provides a measure of what could be gained by adjusting the size of the firm. In a firm, this can shape the strategic planning process and help firms decide whether to choose an expansion or a contraction strategy. For a regulator or researcher, it can indicate the structural efficiency of the industry, i.e. to what extent we have the right number of firms of the right size.

There are, however, some *caveats*. First, the idea of adjusting scale size may not work in reality because the markets may not be competitive and some firms may for natural reasons be unable to change their scale of operation: e.g., if they serve a geographically isolated area of sub-optimal size. We will show how to deal with such complications in Chap. 9. Secondly, the optimal scale size depends on the exact direction in the input and output space. It is therefore not easy to derive simple guidelines on this subject. The optimal size of a farm, for example, can usually not be summarize in a single measure like the amount of acres or the number of cows since it varies with the exact composition of inputs and outputs. A farm specializing in crop production may need to be one size to minimize average costs, while a mixed farm with animals and crops may need to be another size.

Numerical example in R

To illustrate the analysis of scale efficiency, consider the same six firms as in Sect. 4.6.1. We can analyzing their scale efficiencies with the follow R code:

```
> e_vrs <- dea(x,y,RTS='vrs')
> e_drs <- dea(x,y,RTS='drs')
> e_crs <- dea(x,y,RTS='crs')
> se <- eff(e_crs)/eff(e_vrs)
> se
[[1] 0.8000 0.9000 1.000 0.9600 0.6857 0.8000
> abs(eff(e_vrs)- eff(e_drs)) < 1e-4 #test if DRS eff = VRS eff
[1] FALSE FALSE  TRUE FALSE  TRUE FALSE
```

We see that firms A and B are below, firm C is at optimal, firm D is below, firm E is above and F is below optimal scale size.

4.8.2 Allocative efficiency in DEA

In Chap. 2, we also introduced the notion of allocative efficiency as useful for supplementing pure technical efficiency. In terms of input, allocative efficiency AE is related to the choice of the least costly resource mix; in terms of output, it is related to the chosing a revenue-maximizing product mix. It is easy to use these concepts with any specific technology, including the technologies used in DEA. In fact, in DEA models, all resulting optimization problems become simple linear programming problems.

To illustrate this, let w be the input prices. The cost minimal plan is developed by minimizing the cost associated with producing a given output. Thus, as explained in Chap. 2, we must solve the minimization problem*allocative efficiency AE*

$$\min wx \quad \text{subject to} \quad (x, y) \in T$$

Using a VRS DEA technology yields the following LP problem

$$\begin{aligned} \min_{x_1,\ldots,x_m,\lambda^1,\ldots,\lambda^K} \; & w_1x_1 + \cdots + w_mx_m \\ \text{s.t.} \quad x_i &\geq \sum_{k=1}^{K} \lambda^k x_i^k, \qquad i = 1,\ldots,m \\ y_j &\leq \sum_{k=1}^{K} \lambda^k y_j^k, \qquad j = 1,\ldots,n \\ & \sum_{k=1}^{K} \lambda^k = 1. \end{aligned}$$

This is a simple linear programming (LP) problem with $m + K$ variables; we pick m cost-minimizing inputs $x = (x_1, \ldots, x_m)$, and we ensure that they are able to produce y by requiring that there be a convex combination of production plans that (weakly) dominates (x, y).

If we solve the above problem, we find the minimal cost of producing y. We may denote this $C^*(y)$, i.e. $C^*(y)$ is the optimal value of the objective in the above LP problem. Specifically, if we solve it for $y = y^o$, we can therefore find the cost efficiency of firm o as

$$CE(x^o, y^o) = \frac{C^*(y^o)}{wx^o}$$

i.e. as the minimal cost divided by the actual costs. Also, we can find the allocative efficiency as

$$AE(x^o, y^o) = \frac{CE(x^o, y^o)}{E(x^o, y^o; vrs)}$$

In a similar way, if we know the output prices, we can find the maximal revenue production plan by solving a simple LP problem with $n + K$ variables. From this

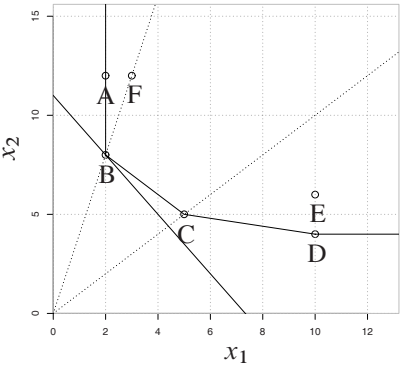
we can evaluate revenue efficiency and hereby also the allocative efficiency on the output side as explained in Chap. 2.

Numerical example in R

Consider the problem with six firms as detailed and illustrated in Table 4.4. We analyzed the same problem in Sect. 2.6.1 using simple graphics and calculation by hand, but we will now show how to do the calculations in R.

Table 4.4 Data for cost minimization

Firm	x_1	x_2	y
A	2	12	1
B	2	8	1
C	5	5	1
D	10	4	1
E	10	6	1
F	3	12	1
Price w	1.5	1.0	



To solve the cost minimization problem using R, we first load the data and then use the procedure `cost.opt` from the Benchmarking package:

```
> library(Benchmark)
> x <- matrix(c(2, 2, 5, 10, 10, 3, 12, 8, 5, 4, 6,12),ncol=2)
> y <- matrix(rep(1,6), ncol=1)
> w <- matrix(c(1.5, 1), ncol=2)
> te <- dea(x,y,RTS="vrs")
> te
[1] 1.0000 1.0000 1.0000 1.0000 0.7500 0.6667
> xopt <- cost.opt(x,y, w, RTS="vrs")
> xopt
     [,1] [,2]
[1,]    2    8
[2,]    2    8
[3,]    2    8
[4,]    2    8
[5,]    2    8
[6,]    2    8
>
> cobs <- x %*% t(w)
> copt <- xopt$xopt %*% t(w)
> ce <- copt/cobs
> ae <- ce/te$eff
```

```
> cbind(ce,te$eff,ae)
          [,1]      [,2]      [,3]
[1,] 0.7333333 1.0000000 0.7333333
[2,] 1.0000000 1.0000000 1.0000000
[3,] 0.8800000 1.0000000 0.8800000
[4,] 0.5789474 1.0000000 0.5789474
[5,] 0.5238095 0.7500000 0.6984127
[6,] 0.6666667 0.6666667 1.0000000
```

We see how the values for technical efficiency (`te`), cost efficiency (`ce`) and allocative efficiency (`ae`) are all similar to those that we calculated by hand in Sect. 2.6.1.

Note also that the `cost.min` method takes the same options as the `dea` method (except, of course, that it also requires a vector with input prices `w`). If the firms have different prices for their inputs, then the input prices must be in the form of a matrix. Because all of the firms have the same input prices and the same output, the optimal input combination is also the same for all firms: $(x_1^*, x_2^*) = (2, 8)$ corresponding to firm B in the figure. To calculate cost efficiency, we find the actual cost wx and the optimal cost wx^*, and we divide the latter by the former. This is calculated in R by using the inner product `%*%`, or matrix multiplication where the function `t` is matrix transposed. Note that firm B is fully technically efficient and displays allocative efficiency; therefore, it is also fully cost efficient. Firms A, C and, D are technically efficient but do not demonstrate allocative efficiency, whereas the reverse is true for firm F, and firm E fails to demonstrate either of these types of efficiency. That F is allocative efficient can also be seen from the figure in Table 4.4; F is on the ray through Firm B, which displays allocative efficiency.

4.9 Summary

In this chapter, we have covered the basic Data Envelopment Analysis (DEA) methods. DEA essentially provides a mathematical programming method for estimating optimal production frontiers and evaluating the relative efficiency of different entities.

The DEA methods can be used to estimate best-practice technologies based on observed production plans using the minimial extrapolation principle. We find the smallest production possibility set that contains data and has a minimum of production economic regularities.

The different DEA methods differ in the ex ante assumptions that they involve. The basic assumptions, as we have also discussed in Chap. 3, are free disposability, convexity and decreasing, increasing or constant returns to scale. Another conceptually appealing assumption is additivity. We have discussed different combinations of these assumptions as they have been made popular in the FDH, VRS, DRS, IRS, CRS and FRH models. We have also shown what the minimal extrapolation technologies look like in these models.

Most DEA studies use is the Farrells notion of efficiency measured as the largest possible proportional contraction of all inputs or the largest possible proportional expansion of all outputs. The combined technology estimation and measurement problem can be formulated as separate simple linear programming problems in the VRS, DRS, IRS and CRS cases. In the FDH and FRH cases, one needs to use more advanced mixed integer programming, although the efficiency of the FDH model can also be determined using simple enumeration techniques.

One of the popular features of DEA models is that they produce explicit peers (i.e. an explicit list of a few firms that a given firm is benchmarked against). The peer units are the firms with positive weights in the evaluation of a given firm. They can guide the learning process and validate the model.

Lastly, we have discussed scale efficiency and allocative efficiency. Scale efficiency is the ability to get the most outputs per input, and it is measured as the ratio of CRS-based efficiency and VRS-based efficiency. Allocative efficiency is to use a cost-minimizing input combination or to produce a revenue-maximizing output mix. The cost-minimizing and revenue-maximizing reference plans can be found by solving simple linear (or, in the FDH and FRH models, mixed integer) programming problems.

In addition to covering the basic DEA models, we have illustrated the use of R and the benchmark library to make actual calculations efficiently.

4.10 Bibliographic notes

DEA was originally proposed by Charnes et al (1978, 1979) and has subsequently been refined and applied in a rapidly increasing number of papers. In his 1992 bibliography, Seiford (1994) lists no fewer than 472 relevant published articles and Ph.D. theses. A 2002 bibliography by Tavaras (2002) includes more than 3000 contributions. For alternative textbook introductions to DEA, see Charnes et al (1995), Coelli et al (1998a), or Cooper et al (2007)

In the DEA literature, the CRS model is often called the CCR model, named after the seminal papers Charnes et al (1978, 1979). The VRS model is often called the BBC model after Banker et al (1984).

Convexity is a strong assumption that is debated in the DEA literature, and different relaxations of the concept have been proposed: e.g., Bogetoft (1996), Bogetoft et al (2000), Chang (1999), Kuosmanen (2001), Petersen (1990), Post (2001), and Tulkens (1993). One reason for the appeal of the convexity assumption in microeconomics is mathematical convenience. With convex sets, prices are useful controls and offer a dual representation based on the idea of separating hyperplanes. Other more basic motivations include

The idea of most productive scale size was suggested by Banker (1984). The concept of using $\sum_k \lambda^k$ to determine firm size as compared to optimal scale size has been discussed by Banker et al (1984), Banker and Thrall (1992), and Chang and Guh (1991), among others.

The R library FEAR for efficiency calculations with R is documented in Wilson (2008). The R package Benchmarking that we generally rely on is introduced in Bogetoft and Otto (2010)

The development of the Swedish DEA models for DSO regulation is described in Agrell and Bogetoft (2000). Using the idea of quasi-observations in the German regulation of transmission companies is discussed in Agrell and Bogetoft (2010a), while the use of management consultants in the regulation in Chile is discussed in Agrell and Bogetoft (2003).

Linear Programming, i.e. linear optimization, is the subject of a large number of standard operations research and mathematical programming books. An early contribution that also emphasize economic applications is Gale (1960). Another early classic on LP and economics models, is Dorfman et al (1958). It is less mathematical. Other old but standard references are Hadley (1962), Hillier and Lieberman (2010), and Luenberger (1984).

The characterization of DEA "cost" and production functions given in the appendix are developed in Bogetoft (1997), Bogetoft (1994b, 1996, 1997), where we also cover some other cases.

4.11 Appendix: More technical material on DEA models

In this appendix, we cover some more technical material relevant to the DEA models. We first prove that the minimal extrapolation principle does lead to the DEA estimated technologies formulated in the main text. Next, we cover a few fundamental results on Linear Programming (LP), and finally, we consider the special cases of a simple cost function and a simple production function in the DEA framework. We derive the properties (increasing, convex/concave, etc) of these functions that corresponds to the properties (free disposability, convexity etc) we have used to characterize the technology set in the main text.

4.11.1 Why the $T^*(\gamma)$ sets work

In this chapter, we have formulated the minimal extrapolation sets T^* that result from a set of K observations when we invoke the assumptions of the different DEA models. To actually prove that $T^*(\gamma)$ is the smallest set containing data and fulfilling the assumptions we have listed for the model referred to as γ, we need to show three things: that the set contains the data, has the desired properties and is the smallest set with these properties.

That the data is included is the easiest to prove. In all cases (i.e. for all the models γ), if $(x^k, y^k) \in T^*(\gamma)$, we can simply pick $\lambda^k = 1, \lambda^{k'} = 0$ for all k' $\neq k$.

To prove that the sets have the stipulated properties is somewhat more complex, or at least, tedious. Consider, for example, the free-disposability property. We must

show that

$$(x, y) \in T^*(\gamma), x' \geq x, y' \leq y \Rightarrow (x', y') \in T^*(\gamma)$$

Now, $(x, y) \in T^*(\gamma)$ means that there exists a $\lambda \in T^*(\gamma)$ such that $x \geq \sum \lambda^k x^k$ and such that $y \leq \sum \lambda^k y^k$. Now, because $x' \geq x$, $y' \leq y$, we can actually use the same λ for $(x'y')$ and we get $x' \geq x \geq \sum \lambda^k x^k$, $y' \leq y \leq \sum \lambda^k y^k$ such that also $(x'y')$ is in $T * (\gamma)$

Lastly, to show that the sets are the smallest sets containing data that have these properties, we shall simply show that at least these sets must be possible because we have already shown that they have the desired properties. Again, this process is a little tedious but straightforward. Note that if the observed practices are feasible, so is

$$(\tilde{x}, \tilde{y}) = (\sum \lambda^k x^k, \sum \lambda^k y^k)$$

for $\lambda \in \Lambda^K(\gamma)$. In the VRS case, this follows directly from the convexity assumption; in the DRS case, it follows from the convexity assumption and decreasing returns to scale (because if we multiple the convex weights by a factor slightly less than 1, then we can simply redefine the weights and let them sum to slightly less than 1). In the IRS case, this follows from the convexity assumption and DRS; in the CRS case, it follows from convexity and CRS. In the FRH case, it follows directly because the right hand side will be one of the observations, and in the FRH case, it follows from possibly repeated use of the additivity condition. Now, because $(\tilde{x}, \tilde{y}) \in T^*(k)$, so is (x, y) with $x \geq \tilde{x}$ and $y \leq \tilde{y}$ based on the free disposability property. This shows that $T^*(\gamma) \subseteq T^*$, and because we have shown that $T^*(k)$ also has the desired properties itself, it must be the minimal extrapolation set.

4.11.2 Linear programming

Linear programming means linear optimization and the general problem is to find a non-negative m-column-vector x which satisfy a system of linear inequalities $Ax \leq b$ for which the linear function cx has a maximum. We can write this as

$$\max_x \; cx \quad \text{subject to} \quad Ax \leq b, \quad x \geq 0$$

where c is a m-row-vector, A a $m \times n$ matrix and b a n-column-vector. The above is often called the *primal LP problem*. We are thus seeking m non-negative variables that maximizes cx subject to n linear restrictions. Without the use of matrices we can formulate this LP problem as

$$\begin{aligned} &\max_{x_1,\ldots,x_m} \; c_1 x_1 + \cdots + c_m x_m \\ &\text{s.t.} \qquad a_{i1} x_1 + \cdots + a_{im} x_m \leq b_i \qquad i = 1, \ldots, n \\ &\qquad\qquad\qquad\qquad x_j \geq 0 \qquad j = 1, \ldots, m \end{aligned}$$

where s.t. is an abbreviation of subject to. We say that a linear program is *feasible* if there exists a vector that satisfy all the linear inequalities. Any such vector is called a *feasible solution*. If there exist an optimal solution x^* we call cx^* for the *value* of the program.

Without further ado we list the main theorems of linear programming that we use in the text. Readers interested in proofs can look in the referred literature or any standard mathematical programming book.

The Fundamental Theorem of LP. *If a LP problem with n restrictions has an optimal solution then there exists an optimal solution in which at most n variables are positive.*

A solution with the mentioned number of positive variable is called a basic solution. The theorem reflects that solutions to LP problems are corner solutions.

An important theme in LP is duality. We will make use of duality in Chap. 5. The dual of the above (primal) LP problem is another LP problem, where the aim is to solve:

$$\min_{y} \; yb \quad \text{subject to} \quad yA \geq c, \quad y \geq 0$$

where y is a n-row-vector, which we typically interpret as prices on the n resources in the b vector. If we look at the primal problem as one of maximizing the value of a production plan, the dual problem is one of finding prices leading to the minimal valuation of the b resources such that no feasible production is profitable.

The Duality Theorem of LP. *If the primal has an optimal solution, x^*, then the dual also has an optimal solution, y^* and vise versa. Moreover, when both have optimal solution, the value of the programs are the same, i.e. $cx^* = y^*b$.*

We can reformulate a LP problem as a Lagrange problem and interpret it as such. The Lagrange function for the primal problem is

$$\mathcal{L}(x, y) = cx + y(b - Ax)$$

where the dual variable y is the Lagrange multipliers. The economic interpretation of a Lagrange multiplier is well-known as the marginal change in the optimal value when we make a marginal change to the side condition, i.e. a marginal change in b. We can also say that y is the shadow price of b. If a condition in optimum is not-binding, i.e. there is a i for which $\sum_{j=1}^{m} a_{ij}x_j < b_i$, then the corresponding optimal dual variable is zero, $y_i = 0$. This result is the equilibrium theorem for LP problems:

The Equlibrium Theorem. *A feasible solution $(x_1, \ldots, x_m)$ for the primal problem and a feasible solution $(y_1, \ldots, y_n)$ for the dual problem are optimal if and only if*

$$y_i = 0 \quad \textit{when} \quad \sum_{j=1}^{m} a_{ij}x_j < b_i \qquad i = 1, \ldots, n$$

and

$$x_j = 0 \quad when \quad \sum_{i=1}^{n} y_i a_{ij} > c_j \qquad j = 1, \ldots, m$$

The conditions are also known as the *complementary slackness conditions*, and the theorem is sometime called The Complementary Slackness Theorem. The first condition conveys a fairly simple economic logic: If there is slack (leftovers) in a constrained primal resource, then additional quantities of that resource must have no value. The second can be given a similar interpretation - if an activity (column in A matrix) is non-profitable, then we should not use it.

4.11.3 DEA "cost" and production functions

The DEA technologies are usually characterized as sets in general input-output space as we have presented them in this chapter. This is useful for general multi–input multi–output settings. When the inputs or the outputs can be aggregated into one dimension (e.g. a cost aggregate or a compound product), we can derive similar representations in function space. Because many practical applications have only one input or one output, it is useful to know which ex ante assumptions the DEA approaches entail in these cases. This is one way to understand the weak ex ante assumptions made under the DEA approach as opposed to those made based on traditional statistical models.

The single input "cost" function

First consider the setting with a single input $m = 1$ interpreted here as costs. We will talk about this as a "cost function". It is a cost function in the sense that it maps outputs into a single input (which we call costs) but deviates from the production economics idea of a cost function as mapping the product of n dimensional output space and m dimensional input price space onto real numbers. On the other hand, the everyday use of the phrase "cost function" is consistent with the situation that we consider here. In regulatory settings, for example, we routinely estimate cost functions linking actual operating expenses to a series of cost drivers reflecting the output services provided. Such cost functions are effectively mappings from n dimensional output space to 1 dimensional input space.

More formally, let there be K observations $(x^k, y^k), k \in K$, associated with an underlying but unknown technology $T \subseteq \mathbb{R}_+^{1+n}$, and let us define a "cost function" or input requirement function as

$$C(y) := \min\{x \mid (x, y) \in T\}$$

The regularities $A1-A4$ that we have used to characterize T in Sect. 4.4 are similar to those for the cost function

$A1^*$ Increasing: $y' \geq y \Rightarrow C(y') \geq C(y)$
$A2^*$ Convex: $C(\alpha y + (1-\alpha)y') \leq \alpha C(y) + (1-\alpha)C(y'), \forall \alpha \in [0,1]$
$A3^*(\gamma)$ $\gamma -$ returns to scale: $C(\kappa y) \leq \kappa C(y), \quad \forall \kappa \in \Gamma(\gamma)$
$A4^*$ Sub-additive : $C(y+y') \leq C(y) + C(y')$

That is, if T satisfies any condition A, $C(\cdot)$ satisfies the corresponding condition A*. The equivalence is actually more involved than that. As long as we have free disposability, we can also construct T from C ,

$$T = \{(x,y) \in \mathbb{R}_+^{1+n} \mid x \geq C(y)\}$$

and the fact that $C(\cdot)$ fulfills any A* now makes T fulfill A. The proofs are not terribly involved, although they are tedious.

It follows that the initial uncertainty about the technology that the different DEA models reflect can be expressed as uncertainty about what the cost function looks like within one of the following broad classes of cost functions:

$$\begin{aligned}
\mathcal{C}(\text{crs}) &= \{C : \mathbb{R}_+^m \to \mathbb{R}_+ \mid C \text{ is increasing, convex, crs}\} \\
\mathcal{C}(\text{drs}) &= \{C : \mathbb{R}_+^m \to \mathbb{R}_+ \mid C \text{ is increasing, convex, drs}\} \\
\mathcal{C}(\text{vrs}) &= \{C : \mathbb{R}_+^m \to \mathbb{R}_+ \mid C \text{ is increasing, convex}\} \\
\mathcal{C}(\text{frh}) &= \{C : \mathbb{R}_+^m \to \mathbb{R}_+ \mid C \text{ is increasing, sub-additive}\} \\
\mathcal{C}(\text{fdh}) &= \{C : \mathbb{R}_+^m \to \mathbb{R}_+ \mid C \text{ is increasing}\}
\end{aligned}$$

Thus, in the DEA framework, there is considerable a priori uncertainty. In a γ-DEA model, we know that

$$C(\cdot) \in \mathcal{C}(\gamma)$$

i.e. we know that $C(\cdot)$ is increasing when γ = fdh, for example, but otherwise, we know nothing about the cost function. These results also emphasize why the DEA approach has been termed non-parametric. Our a priori uncertainty does not stem from a lack of information about a few parameters, as in a Coob-Douglas or a Translog statistical model. Rather, we lack information about all of the characteristics of the function except for a few general properties such as its tendency to increase.

The *minimal extrapolation principle* for estimating T as the smallest set T^* containing data and satisfying the conditions imposed can now be translated into function space. We estimate the DEA-based cost function as the largest function with these properties that is consistent with data in the sense that $x^k \geq C(y^k)$, $k = 1, \ldots, K$, .i.e. as

$$C^*(y) := \max\{C(y) \mid C(\cdot) \in \mathcal{C}(k), x^k \geq C(y^k) \quad k = 1, \ldots, K\}$$

We see that this approach involves an *outer approximation* of the production possibility set $T = \{(x, y)|x \geq C(y)\}$ as opposed to the inner approximation we have worked with above. The approximation is illustrated in the left panel of Fig. 4.5 below for a DRS technology.

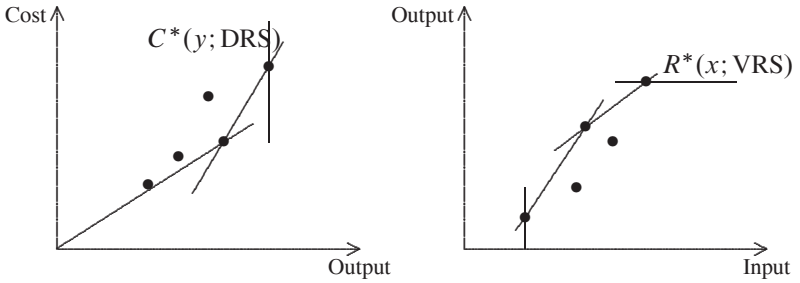


Fig. 4.5 DEA cost and production function estimations

The cost function approximation concept is an important alternative view of the DEA cost function. Assume that a principal bargains with an agent regarding the production of some output y. $C^*(y)$ can be interpreted as the highest cost that the agent can possibly claim when it is publicly known that x^k can produce y^k for all $k = 1, \ldots, K$. If the principal has no information on any subject other than the type of cost function and the K observations, he cannot pay less than $C^*(y)$ if he wants to be sure that the agent will accept the contract because there is always a possibility that the true cost does coincide with $C^*(y)$. Hence, paying anything less than $C^*(y)$ will involve the risk that the agent will quit the relationship.

An important feature of the class of cost functions considered in DEA is that a largest function of this type actually exists–or, to put it differently, that the $C^*(\cdot)$ defined above inherits the properties from the $\mathcal{C}(k)$ class. This is the functional version of the discussion that we had above about which assumptions one can combine with the idea of minimal extrapolation.

Now the largest cost function that is consistent with data and has the desired regularity, $C^*(\cdot)$, is also the cost function that we would obtain by first doing minimal extrapolation in the input-output space (i.e. by first constructing T^* and then deriving the cost function). That is, we have $C^*(\cdot) = C^{**}(\cdot)$ where

$$\begin{aligned} C^{**}(y) = \min_{x,\lambda}\ & x \\ \text{s.t.}\quad & x \geq \textstyle\sum_k \lambda^k x^k \\ & y \leq \textstyle\sum_k \lambda^k y^k \\ & \lambda \in \Lambda^K(\gamma), \quad x \in \mathbb{R}_+ \end{aligned}$$

This means that even though we use an outer approximation strategy, we end up with the same estimated function as when we use an inner approximation strategy on T.

In summary, in the single-input, multiple-output case, we may think of DEA in the usual way using production sets with desired properties, or we may think of it in cost function space where the cost function features related and natural regularities.

The single-output production function

The same equivalence can of course be established in the multi–input, single-output case $n = 1$, where we can think of the production possibilities in terms of traditional production functions. When $T \subseteq \mathbb{R}_+^{m+1}$ is the production possibility set, we define the production function in the usual way as the maximum possible output from a given input

$$R(x) := \max\{y \mid (x, y) \in T\}$$

The regularities $A1 - A4$ that we have used to characterize T translate into similar regularities of the production function

$A1^{**}$ Increasing : $x' \geq x \Rightarrow \pi(x') \geq \pi(x)$
$A2^{**}$ Concave : $\pi(\alpha x + (1-\alpha)x') \geq \alpha\pi(x) + (1-\alpha)\pi(x'), \forall \alpha \in [0, 1]$
$A3^{**}(\gamma)$ $k-$ returns to scale : $\pi(\kappa x) \geq \kappa\pi(x), \quad \forall \kappa \in \Gamma(\gamma)$
$A4^{**}$ Super-additive : $\pi(x + x') \leq \pi(x) + \pi(x')$

That is, if T satisfies any condition A, R satisfies the corresponding condition A** and vice versa, as in the cost case. The classes of production functions corresponding to the classical DEA models are now

$$\begin{aligned}
\mathcal{R}(\text{crs}) &= \{R \mid R \geq 0 \text{ increasing, concave, crs}\} \\
\mathcal{R}(\text{drs}) &= \{R \mid R \geq 0 \text{ increasing, concave, drs}\} \\
\mathcal{R}(\text{vrs}) &= \{R \mid R \geq 0 \text{ increasing, concave}\} \\
\mathcal{R}(\text{fdh}) &= \{R \mid R \geq 0 \text{ increasing}\} \\
\mathcal{R}(\text{frh}) &= \{R \mid R \geq 0 \text{ super-additive}\}
\end{aligned}$$

That is, an alternative interpretation of the DEA framework in the single-output case is that we know ex ante that the underlying production function belongs to a certain wide class of functions

$$R(\cdot) \in \mathcal{R}(\gamma)$$

but that we otherwise have no a priori information about which production function prevails.

The *minimal extrapolation principle* in the production function representation estimates the smallest function with these properties that is consistent with data in the sense that $R(x^k) \geq y^k$, $k = 1, \ldots, K$,

$$R^*(x) := \min\{R(x) \mid R(\cdot) \in \mathcal{R}(k), \quad y^k \leq R(x^k) \quad k = 1, \ldots, K\}$$

Conceptually, the idea is that the firms may have lost some of their potential output and may therefore be working below the frontier. Minimal extrapolation approximation minimizes the potential productions and thereby the contemplated losses that come as a result of being off the frontier.

We see that the production function approach, like the cost function approach, involves the *outer approximation* of the production possibility set T rather than the inner approximation (as in the models we used in the main text). The resulting approximations are the same as long as we choose the classes of production functions appropriately. The approximation is illustrated in the right panel of Fig. 4.5 above for a VRS technology.

Finally, we note that the lowest production function that is consistent with the data and has the desired regularity, $R^*(\cdot)$, is also the production function that we would obtain by first doing minimal extrapolation in the input-output space (i.e. by first constructing T^* and then deriving the production function from this). That is, we have $R^*(\cdot) = R^{**}(\cdot)$ where

$$\begin{aligned} R^{**}(x) = \max_{y,\lambda}\ & y \\ \text{s.t.}\quad & x \geq \textstyle\sum_k \lambda^k x^k \\ & y \leq \textstyle\sum_k \lambda^k y^k \\ & \lambda \in \Lambda^K(\gamma),\ y \in \mathbb{R}_+ \end{aligned}$$

In summary, in the single output case, we may think of DEA in the usual way using production sets with desired properties, or we may think of it in production function space where the production function has related and natural regularities.

Chapter 5
Additional Topics in DEA

5.1 Introduction

We covered the basics of DEA in the previous chapter. In this chapter, we will cover some additional topics that we find particularly relevant both to applications and to our understanding of DEA.

We will first discuss the idea of super-efficiency. Next, we will relax the Farrell idea of proportional improvements in all inputs or outputs. We will discuss situations in which only some of the resources or services are discretionary, and more generally, we will discuss the use of directional distance measures in DEA. We will close our discussion of alternative efficiency concepts in the DEA context by taking a closer look at the slack problem.

The last half of this chapter considers more radical reformulations of the basic problems. We will look at dual versions of our traditional DEA programs and consider how they can be used to provide alternative interpretations of DEA measures and to impose restrictions on the relative importance of different resources and services. We will also discuss some minimax formulations and explain why such game formulations can be useful. In the Appendix, we discuss different types of outliers and how to identify and eliminate outliers in a frontier model.

Statistical inference in DEA models is covered in the next chapter.

5.2 Super-efficiency

What is now called super-efficiency and is routinely calculated using several software programs was first suggested as a means of differentiating among frontier units. In many applications, several firms are ranked as fully efficient, and it may be interesting to consider ways of ranking them.

The idea of super-efficiency was later proved crucial to regulation and contracting applications of DEA. It is intuitively obvious that firms with an efficiency score of

P. Bogetoft and L. Otto, *Benchmarking with DEA, SFA, and R*, International Series in Operations Research & Management Science 157, DOI 10.1007/978-1-4419-7961-2_5,

1 have little incentive to improve because it will not improve their score. We will return to this in more detail in Chap. 10.

Super-efficiency measures are constructed by avoiding that the evaluated firm can help span the technology. Let $T^*(\gamma \mid -k)$ be a DEA approximation of the technology using the γ assumptions and based on all observations but that of firm k:

$$T^*(\gamma \mid -k) = \left\{ (x, y) \in \mathbb{R}^m_+ \times \mathbb{R}^n_+ \mid \exists \lambda \in \Lambda^{K-1}(\gamma) : x \geq \sum_{j \neq k} \lambda^j x^j,\ y \leq \sum_{j \neq k} \lambda^j y^j \right\}.$$

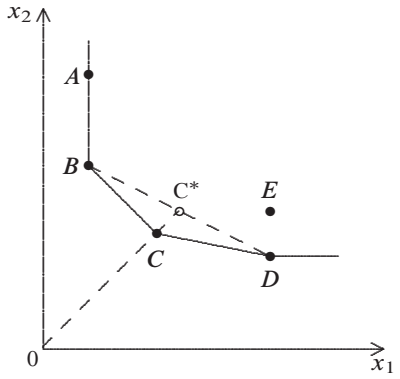
Now the efficiency of (x^k, y^k) relative to $T^*(\gamma \mid -k)$ is called *super-efficiency*

$$E^{\text{SUP}\,k} = E\big((x^k, y^k);\ T^*(\gamma \mid -k)\big),$$
$$F^{\text{SUP}\,k} = F\big((x^k, y^k);\ T^*(\gamma \mid -k)\big).$$

Consider the data in Table 5.1. Assuming VRS, the input isoquant is graphed in the figure to the right of the table. Based on the figure, we can see that the usual

Table 5.1 Super efficiency

Firm	x_1	x_2	y
A	2	12	1
B	2	8	1
C	5	5	1
D	10	4	1
E	10	6	1



input efficiency of firm C is 1, $E^1 = 1$. If we leave C out of the set of firms that generate the technology, then the isoquant corresponds to the dashed line. The super-efficiency of firm C must be evaluated against this isoquant instead of the original one, and we obtain $E^{\text{SUP}\,C} = 1.2$. The corresponding projection of C becomes C^*.

It is simple to set up the associated mathematical programs—they look just like the usual ones except that one column has been eliminated that corresponds to the λ^k variable.

The super-efficiency measures on the input and output sides are not restricted to either below or above 1. Indeed, this is part of the motivation for using them—we are interested in differentiating among the firms with traditional efficiency scores of 1. The input super-efficiency score $E^{\text{SUP}\,k}$ may be larger than 1; firm k could

have increased its inputs by a factor $E^{\text{SUP}\,k}$ and still not have been dominated by a feasible reference unit. Likewise, $F^{\text{SUP}\,k}$ can be smaller than 1 if firm k could have reduced all of its outputs by a factor $F^{\text{SUPER}k}$ without being dominated by a reference unit.

It also follows from the definition that the traditional efficiency measures are simply aggregates of the super-efficiency measures

$$E^k = \min\{E^{\text{SUP}\,k}, 1\} \quad \text{and} \quad F^k = \max\{F^{\text{SUP}\,k}, 1\}.$$

Hence, the super-efficiency measures contain at least the same information and sometimes also contain additional information. It is therefore obvious that they are advantageous for decision-making and incentive purposes—at least as long as we ignore possible information-processing costs.

The only drawback of the Farrell-based super-efficiency measures is that the resulting programs may not have feasible solutions. In those cases, we define the super-efficiencies as

$$E^{\text{SUP}\,k} = \infty \quad \text{and} \quad F^{\text{SUP}\,k} = -\infty$$

respectively. The presence of infinite super-efficiencies simply means that there are no other units against which to gauge firm k with the given data and the imposed technological regularities. Such firms are sometimes referred to as *hyper-efficiency*. Using the usual Farrell measures, we can always find solutions to the LP-problems and hyper-efficient firms would have be classified as fully efficient.

Mathematically, these definitions are natural because they correspond to inf and sup over empty sets, meaning that min and max do not exist. What is important, however, is the conceptual idea. Sometimes, we may not be able to find reference firms to gauge the evaluated firm against, in which case we effectively put the evaluated firm in the best possible light using these definitions.

This is not to say that the existence problem is without practical relevance. In practice, and particularly when we base decision-making and incentive procedures on super-efficiency, the lack of solutions does create some complications in the sense that the corresponding firms need special care. In a regulatory context, for example, the extremely super-efficient units may be transferred to individual evaluation by a regulator who has otherwise mechanized his decision-making. Of course, there are many other ways to avoid such problems, and we will discuss some of these as we consider more practical implementations of DEA-based incentive schemes. For now, let us just mention a few more technical solutions. We could eliminate or reduce the problem of infinite efficiencies by introducing a) other technological assumptions (e.g., more re-scaling possibilities); b) full or partial aggregations of some of the inputs and outputs (e.g., by using partial information about costs and price elements, cf. below); and c) supplementary observation (e.g,. engineering phantom observation used to supplement the observed best practices, cf. eg. Sect. 4.7).

Let us also emphasize that the idea of super-efficiency is not solely associated with the Farrell measure. For other measures, including those introduced next, we

can calculate super-efficiency in a similar way by avoiding that the evaluated firm affects the technology against which it is gauged, i.e. by *ensuring that a unit cannot affect its own benchmark*.

Numerical example in R

In R, we can calculate super-efficiency using the `dea` command where we distinguish between the firms that we wish to evaluate and the firms that define the technology. To further facilitate this process, the Benchmarking package includes the function `sdea` in addition to `dea`. The function `sdea` calculates super-efficiencies directly and includes the usual options.

Consider, for example, the data in Table 4.2 on page 95, and let us calculate super-efficiencies using R.

```
> # Input super efficiency, vrs and crs
> sdea(x,y, RTS="vrs", ORIENTATION="in")
[1] 2.0000 0.6667 1.4375 0.5556        NA 0.4000
> sdea(x,y, RTS="crs", ORIENTATION="in")
[1] 0.8000 0.6000 1.2500 0.5333 0.6857 0.3200
> # Output super efficiency, vrs and crs
> sdea(x,y,RTS="vrs",ORIENTATION="out")
[1] NA 1.6667 0.7200 1.4167 0.8333 2.6667
> sdea(x,y,RTS="crs",ORIENTATION="out")
[1] 1.250 1.667 0.800 1.875 1.458 3.125
```

Note that the input super-efficiency in the VRS model for firm E has the value `NA`, showing that there is no solution to the DEA program. This can also be seen in Fig. 4.2 in which the observation for firm E is not in the technology set if E is not included in the data set that generates the technology set; we can see that a horizontal line through E never intersects the VRS technology set without E. In other words, if we believe that the VRS technology set is valid, it seems that firm E could expand its input consumption infinitely without becoming inefficient. Similarly, firm A could reduce its outputs infinitely without becoming inefficient in the VRS technology set.

5.3 Non-discretionary variables

The DEA models considered so far all rely on the Farrell approach to efficiency measurement; all inputs are reduced or all outputs are expanded by the same factor. This type of proportional adjustment is challenged by a series of alternative efficiency measurement approaches that may be more useful for studying organizational learning, coordination and motivation. We cover a few of these in this section and in upcoming sections. To simplify the exposition, we consider only input efficiency. Parallel treatments of output, however, are straightforward.

An early suggestion is to consider the flexibility of the resources. In some settings, certain improvements may be impossible. In extreme cases, the firm may only

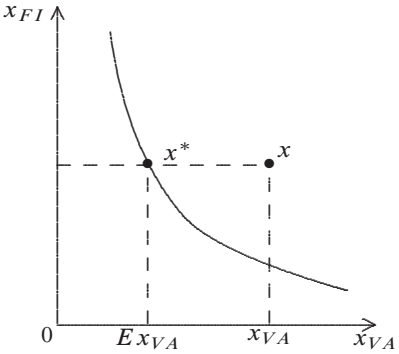


Fig. 5.1 Input efficiency E with fixed input

control some if its inputs, say the set VA of *variable or discretionary inputs*, $VA \subset \{1, \ldots, m\}$. Others, the *fixed, non-discretionary inputs* $FI = \{1, 2, \ldots, K\} \backslash VA = \{ h \in \{1, \ldots, m\} \mid h \notin VA \}$ cannot be adjusted—at least not at the level of the firm in which our production units operate or with the time horizon that we study. Let $x = (x_{VA}, x_{FI})$ denote the variable and fixed inputs.

In such cases, a traditional and popular variation of the Farrell procedure is to look for the largest proportional reduction in the variable inputs alone

$$E\left((x^{\mathrm{o}}_{VA}, x^{\mathrm{o}}_{FI}, y^{\mathrm{o}}); T\right) = \min_{E}\{ E \mid (Ex^{\mathrm{o}}_{VA}, x^{\mathrm{o}}_{FI}, y^{\mathrm{o}}) \in T \}.$$

This leads to simple modifications of the DEA program in which we only reduce in the input rows where the inputs are considered to be variable.

$$\begin{aligned}
\min_{E,\lambda^1,\ldots,\lambda^K} \quad & E \\
\text{s.t.} \quad & Ex^{\mathrm{o}}_i \geq \sum_{k=1}^{K} \lambda^k x^k_i, && i \in VA \\
& x^{\mathrm{o}}_i \geq \sum_{k=1}^{K} \lambda^k x^k_i, && i \in FI \\
& y^{\mathrm{o}} \leq \sum_{k=1}^{K} \lambda^k y^k \\
& \lambda \in \Lambda^K(\gamma)
\end{aligned}$$

We see that the DEA models that distinguish between discretionary (variable) and non-discretionary (fixed) inputs lead once again to simple linear or mixed integer programming problems. This variant of the Farrell measure is illustrated in Fig. 5.1.

Note that the above DEA program can be rewritten as

$$
\begin{aligned}
\min_{E,\lambda^1,\ldots,\lambda^K} \quad & E \\
\text{s.t.} \quad & E x_i^{\mathrm{o}} \geq \sum_{k=1}^{K} \lambda^k x_i^k, && i \in VA \\
& -x_i^{\mathrm{o}} \leq \sum_{k=1}^{K} \lambda^k (-x_i^k), && i \in FI \\
& y^{\mathrm{o}} \leq \sum_{k=1}^{K} \lambda^k y^k \\
& \lambda \in \Lambda^k(\gamma)
\end{aligned}
$$

Hence, the fixed, non-discretionary input corresponds to a negative output. Thus, if we have a problem with fixed inputs, we can still use the usual programs for DEA models if we simply treat the fixed inputs as negative outputs. We should note, however, that most software solutions assume inputs and outputs to be positive. The R library Benchmarking, however, does not restrict the sign of the inputs and outputs.

The approach that involves letting some variables be discretionary and some be non-discretionary is also sometime referred to as a *sub-vector efficiency approach*.

Practical application: Fishery

In an analysis of the efficiency of Danish fishery, we used a representative sample of some 288 Danish fishing vessels.

On the output side, we aggregated the available catch data into nine output groups defined as follows: (1) cod, (2) other gadoids, (3) plaice, (4) other flat-fish, (5) herring, (6) mackerel, (7) lobster and shrimp, (8) other consumption species and (9) industrial species. On the input side, all costs in the dataset were categorized as either variable (discretionary) or fixed (non-discretionary).

The variable costs are expenses for (1) fuel and lubricants, (2) ice and provisions, (3) landings and sales and (4) the crew, whereas the fixed costs include that of (1) maintenance and (2) insurance and various services. Note that because DEA models are unaffected by linear transformations of a given variable, the insurance costs can also be a proxy for the capital costs.

In addition to evaluating individual vessels, we also investigated the impact of reallocation catch values for the different vessels. We will return to some this idea in Chap. 9.

5.4 Directional efficiency measures

One of the ways to generalize the methods used so far is to look for improvements in some arbitrary direction $d \in \mathbb{R}^m_+$. This is the idea behind the directional distance function approach discussed in Chap. 2.

Recall that efficiency in this approach is measured in an additive rather than a multiplicative manner. We measure the distance to the frontier in d-units leading to a directional distance or *excess function*

$$e = e(x^o, y^o;\ T, d) := \max\{\, e \in \mathbb{R}_+ \mid (x^o - ed, y^o) \in T \,\}.$$

The excess $e(x^o, y^o;\ T, d)$ is the number of times the input bundle d has been used in x^o in excess of what is necessary to produce y^o. Hence, a large degree of excess reflects a large (absolute) amount of slack and a considerable amount of inefficiency.

As usual, finding the directional distance or excess in a setting in which the technology is estimated using DEA involves solving a linear or mixed integer program:

$$\begin{aligned} \max_{e,\lambda^1,\ldots,\lambda^K} \quad & e \\ \text{s.t.} \quad x^o - ed &\geq \sum_{k=1}^{K} \lambda^k x^k \\ y^o &\leq \sum_{k=1}^{K} \lambda^k y^k \\ \lambda &\in \Lambda^K(\gamma) \end{aligned}$$

A crucial question that emerges when we use directional distance is *which direction to choose*. We mention four approaches here.

One is to choose $d = x^o$, i.e. to look at improvements in the direction of the actual input consumption. In this case, we obtain

$$E\big((x^o, y^o); T\big) = 1 - e(x^o, y^o; T, x^o).$$

In this sense, then, the traditional Farrell measure is a special instance of the directional distance measure

A second approach is to choose $d = (1, \ldots, 1, 0, \ldots, 0)$ such that the last part of the input–vector is fixed. This is similar to the approach used in the last section, in which we only introduced improvements to the discretionary variables.

A third method is to think of the choice of direction from the point of view of a user—the firm or a principal or regulator—that has particular preferences regarding the inputs.

A fourth way to guide the choice of direction is to think of efficiency improvements as a bargaining process. The different input factors can be thought of as production factors (e.g. different labor types), all of which seek to have excess amounts

available so as to avoid having to invest too much effort and to gain access to the benefits of slack at the workplace. This idea can be formalized along the lines of what has been called *potential improvements* PI or *multi-directional efficiency analysis* MEA. The approach is based on axiomatic bargaining theory analogous to the Kalai and Smorodinsky solution to a two-person bargaining problem and depends not only on the actual consumption mix but also on the shape of the technology in the neighborhood of actual production. This seems a natural property in the sense that if there is significant opportunity to reduce one input and less of an opportunity to reduce another input, then the direction of improvement should lean more toward the first input and less toward the second.

To implement this idea, we first examine the possible savings in the individual inputs presuming that the other inputs are not reduced. That is, for firm o, we calculate the minimal usage of input i when all other inputs are held fixed as

$$\hat{x}_i^{\mathrm{o}} = \min_{x_i}\{\, x_i \mid (x_i, x_{-i}^{\mathrm{o}}, y^{\mathrm{o}}) \in T \,\}$$

where $x_{-i} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_m)$ is the vector where the ith coordinate is missing. Next, the $\hat{x}_i^{\mathrm{o}}$ values are combined into an *ideal* for firm o, $\hat{x}^{\mathrm{o}} = (\hat{x}_1^{\mathrm{o}}, \ldots, \hat{x}_m^{\mathrm{o}})$. This ideal is generally outside the production set; i.e., usually $(\hat{x}^{\mathrm{o}}, y^{\mathrm{o}}) \notin T$. Movement in the direction of this ideal is possible, however, and represents the potential improvement direction

$$x^{\mathrm{o}} - \hat{x}^{\mathrm{o}} = (x_1^{\mathrm{o}} - \hat{x}_1^{\mathrm{o}}, \ldots, x_m^{\mathrm{o}} - \hat{x}_m^{\mathrm{o}})$$

that can now be used to calculate the excess. In addition, the direction can also be interpreted as indicating the improvement potential associated with the different dimensions. In applications, such multi-directional efficiency evaluations have proved to provide useful, nuanced information. In a hospital with twice as many nurses as doctors, for example, it might be interesting to know that there is limited opportunity to reduce the number of nurses but a real opportunity to reduce the number of doctors.

The approach and the hospital example are illustrated in Fig. 5.2, where the dotted line corresponds to the Farrell direction and the dashed line is the potential improvement direction. To determine the potential improvement direction for a given firm o, we solve m linear or mixed integer programming problems, one for each of the inputs. For input h, the program to determine $\hat{x}_h^{\mathrm{o}}$ looks like this:

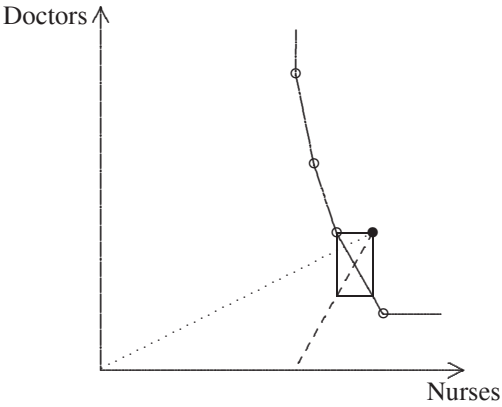


Fig. 5.2 Potential improvement direction

$$
\begin{aligned}
\min_{x_h,\lambda^1,\ldots,\lambda^K} \quad & x_h \\
\text{s.t.} \quad & x_h \geq \sum_{k=1}^{K} \lambda^k x_h^k \\
& x_i^{\mathrm{o}} \geq \sum_{k=1}^{K} \lambda^k x_i^k, \qquad i = 1,\ldots,h-1,h+1,\ldots,m \\
& y_j^{\mathrm{o}} \leq \sum_{k=1}^{K} \lambda^k y_j^k, \qquad j = 1,\ldots,n \\
& \lambda \in \Lambda^K(\gamma)
\end{aligned}
$$

In addition, to calculate the excess for firm o in the potential improvement direction

$$e^{\mathrm{o}} = \max\{\, e > 0 \mid x^{\mathrm{o}} - e(x^{\mathrm{o}} - \hat{x}^{\mathrm{o}}) \in T \,\},$$

we must solve an additional linear or mixed integer problem, as on page 121, where $d = x^{\mathrm{o}} - \hat{x}^{\mathrm{o}}$. In this model, then, excess resources or slack is estimated as

$$z_i^{\mathrm{o}} = x_i^{\mathrm{o}} - \left(x_i^{\mathrm{o}} - e^k(x_i^{\mathrm{o}} - \hat{x}_i^{\mathrm{o}})\right) = e^{\mathrm{o}}(x_i^{\mathrm{o}} - \hat{x}_i^{\mathrm{o}}) \geq 0, \qquad i = 1,\ldots,m$$

which can also indicate strategies for improving performance.

Practical application: Bank branches

An application of MEA involved staffing decisions in 291 branches of a large Canadian bank. The bank has well-developed staffing models, and the branches work in a highly competitive environment. One would therefore expect limited 'inefficiency'

Table 5.2 Relative saving potentials in bank branches

Inputs	Mean	Max
Teller	18%	48%
Typing	47%	91%
Accounting & Ledgers	24%	56%
Supervision	35%	76%
Credit	26%	67%

in the sense of wasted resources and over-staffing. Using DEA, we nevertheless found considerable 'inefficiency'.

The model included 5 different inputs corresponding to full-time positions in various employee groups: Tellers, Typing, Accounting & Ledgers, Supervision, and Credit. The model also included 9 outputs: Counter Transactions, Counter Sales, Security Transactions, Deposit Sales, Personal Loan Sales, Commercial Loans, Term Accounts, Personal Loan Accounts and Commercial Loan Accounts.

As part of the analysis, we compared the staffing profiles developed with the ideal staffing profile as discussed above. For each branch, we therefore calculated the sub-vector efficiency of each of the staff groups separately; i.e., we calculated the relative savings that would accrue if only the number of tellers were adjusted, if only the number of typists were adjusted, and so on. The average and maximum values for the branches are given in Table 5.2 (the minimum values are obviously 0).

The natural question is if this inefficiency is best interpreted as waste or if the apparent inefficiency may serve other purposes. To investigate this question, we invoked the theoretical framework of *rational inefficiency*. Indeed, as the table illustrates, a systematic pattern of slack consumption emerges, suggesting that the allocation of slack between staff groups is far from random.

The systematic pattern seems natural from the point of view of employee value and hierarchy and also when considering employee flexibility and substitutability. Thus, for example, we find a relatively high level of over-staffing at the supervisor level, which is natural given both the strong bargaining position of these individuals based on their role in the branch hierarchy and the relative flexibility of supervisor resources. It therefore appears that the location of the branches in production space is in accordance with the predictions of the rational inefficiency hypothesis, which suggest that slack is not allocated randomly but is rather distributed according to organization preferences, bargaining power, etc.

5.5 Improving both inputs and outputs

As explained in Chap. 2, we can also combine the ideas of input and output efficiency by examining to what extent we can simultaneously use less input and produce more output. Using the direction distance function approach, we can look for

changes in the direction $(d_x, d_y) \in \mathbb{R}^m_+ \times \mathbb{R}^n_+$ and define the directional efficiency e^{o} of firm o as

$$e^{\mathrm{o}} = \max\{\, e > 0 \mid (x^{\mathrm{o}} - e d_x, y^{\mathrm{o}} + e d_y) \in T \,\}.$$

as illustrated in Fig. 2.7.

The corresponding optimization problem when we use the DEA-estimated technology to evaluate (x, y) is

$$\begin{aligned} \max_{e,\lambda^1,\ldots,\lambda^K} \quad & e \\ \text{s.t.} \quad & x^{\mathrm{o}} - e d_x \geq \sum_{k=1}^{K} \lambda^k x^k \\ & y^{\mathrm{o}} + e d_y \leq \sum_{k=1}^{K} \lambda^k y^k \\ & \lambda \in \Lambda(\gamma) \end{aligned}$$

Again, this is a simple linear or mixed integer problem.

A slightly more complicated concept is that of graph or hyperbolic efficiency, as discussed in Chap. 2. Recall that the *graph hyperbolic efficiency measure* Gmeasure of technical efficiency G for firm o is defined as

$$G = \min\{\, G > 0 \mid (G x^{\mathrm{o}}, \frac{1}{G} y^{\mathrm{o}}) \in T \,\}.$$

Inserting the DEA technology, we obtain

$$\begin{aligned} \min_{G,\lambda^1,\ldots,\lambda^K} \quad & G \\ \text{s.t.} \quad & G x^{\mathrm{o}} \geq \sum_{k=1}^{K} \lambda^k x^k \\ & \frac{1}{G} y^{\mathrm{o}} \leq \sum_{k=1}^{K} \lambda^k y^k \\ & \lambda \in \Lambda^K(\gamma) \end{aligned}$$

$$\begin{aligned} \min_{G,\lambda^1,\ldots,\lambda^K} \quad & G \\ \text{s.t.} \quad & Gx_i^{\text{o}} \geq \sum_{k=1}^{K} \lambda^k x_i^k, \qquad i = 1,\ldots,m \\ & \frac{1}{G} y_j^{\text{o}} \leq \sum_{k=1}^{K} \lambda^k y_j^k, \qquad j = 1,\ldots,n \\ & \lambda \in \Lambda^K(\gamma) \end{aligned}$$

We see that even in cases with a DEA-estimated technology, G is a solution to a nonlinear programming problem; we reduce the inputs and expand the outputs by factors G and $\frac{1}{G}$, respectively. Hence, the graph efficiency is slightly harder to find than the other efficiencies we have defined because they all lead to simple LP or mixed integer problems.

In connection with super-efficiency, the graph orientation offers the advantage that there is always a solution to the non-linear programming problem, whereas with the input and output orientation, there might not always be a solution.

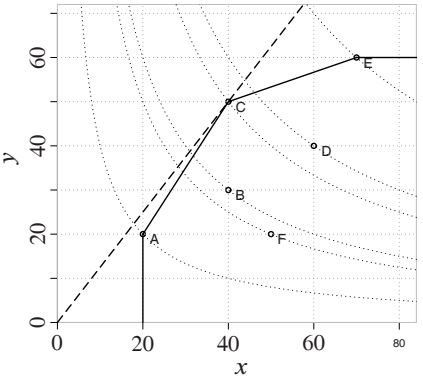
Numerical example in R

We use the data in Table 4.2 on page 95 to calculate the graph efficiency. The calculated efficiencies are shown in Table 5.3, where the technology set with the relevant graph hyperbola curves embedded is also shown.

Table 5.3 Graph efficiency for six firms

Firm	x	y	G^{vrs}	G^{crs}
A	20	20	1.00	0.89
B	40	30	0.80	0.77
C	40	50	1.00	1.00
D	60	40	0.77	0.73
E	70	60	1.00	0.83
F	50	20	0.59	0.57

Note: Compare with Table 4.3 page 97



The calculations are done in R with the `dea` function from the Benchmarking package by using the option `ORIENTATION="graph"` as shown in the following lines of code:

```
> x <- matrix(c(20, 40, 40, 60, 70, 50),ncol=1)
> y <- matrix(c(20, 30, 50, 40, 60, 20),ncol=1)
> dea(x,y,RTS="vrs",ORIENTATION="graph")
[1] 1.0000 0.7953 1.0000 0.7687 1.0000 0.5874
> dea(x,y,RTS="crs",ORIENTATION="graph")
[1] 0.8945 0.7746 1.0000 0.7303 0.8281 0.5657
```

5.6 Slack considerations

One of the drawbacks of the traditional Farrell approach is that a firm can have an efficiency score of 1 and still be Koopmans inefficient in the sense that some inputs could be reduced or some outputs could be expanded without affecting the need for other inputs or the production of other outputs. This is called lack of indication in axiomatic theory, cf. page 54.

This phenomenon is not only theoretically possible but is also rather common in many DEA models and is associated with firms being projected on the vertical or horizontal parts of the production frontier. For example, consider firm A in the data set in Fig. 5.1 on page 116. The input efficiency is $E^A = 1$ because we cannot reduce both inputs proportionally and still be on the isoquant. However, it is obvious from the graph that A is wasting input 2. We say that there is slack in input 2. We can, in fact, reduce input 2 by 4 units without reducing output.

The slack problem has been clear since DEA was first developed, and two principal solutions have been suggested. One is to penalize such slack using what is known as an infinitesimal penalty factor, a factor that is large enough to recognize the possible slack and small enough not to impact the numerical results. We will formalize this method below. The other approach is to solve the dual problem using strictly positive input and output prices. We will discuss the dual problem in the next section. Technically, these two solutions are actually equivalent, or dual, as we say in linear programming.

To penalize slack, we can consider the following reformulation of the Farrell input efficiency program Eq. (4.1) on page 90

$$\min_{E,\lambda} \quad E - \delta \sum_{i=1}^{m} z_i^- - \delta \sum_{j=1}^{n} z_j^+ \tag{5.1}$$

$$\text{s.t.} \quad E x_i^{\text{o}} + z_i^- = \sum_{k=1}^{K} \lambda^k x_i^k, \qquad i = 1, \ldots, m$$

$$y_j^{\text{o}} - z_j^+ = \sum_{k=1}^{K} \lambda^k y_j^k, \qquad j = 1, \ldots, n$$

$$\lambda \in \Lambda^k(\gamma), \quad z^- \geq 0, \quad z^+ \geq 0, \quad E \leq 1.$$

Here we have introduced input slack variables $z_i^-, i = 1, \ldots, m$ that measure any excess resources in Ex^{o} as compared to the reference unit. Likewise, we have introduced output slack variables $z_j^+, j = 1, \ldots, n$ that measure any excess output in the reference unit as compared to firm o. Finally, we have introduced $\delta > 0$ as a penalty for slack. For $\delta = 0$, we return to our original program.

We see that we now have two concerns - we seek to minimize E and simultaneously maximize the slacks. The trade-off between the two depends on δ. The larger δ is, the more we focus on the maximization of slack. It is not surprising that this double concern will generally yield different results. If δ is sufficiently large, we may just want to choose E=1 and then minimize the objective function by having a lot of slack. Hence, what we are ideally looking for is a value of δ that is sufficient small not to impact the choice of E but sufficiently large to create the maximum possible slack. This is the idea of δ being infinitesimal.

The best way to conceptualize the above problem is therefore in *lexicographic* terms. We first minimize E; then, having done this, we maximize the sum of the slack values for the fixed value of E. This approach is known as a *two-stage approach*. We first conduct an ordinary Farrell input efficiency analysis, after which we calculate the maximal slacks given the calculated efficiency level. If there is positive slack, we will say that the firm is Farrell efficient but that there is additional saving potential associated with some inputs and/or the opportunity for expansion associated with some outputs.

One way to look at this two-stage procedure is to *select a good reference firm*. A projection that includes possible slack will not be entirely convincing in the real world. Imagine that such projections are presented to a client and that he is somehow able to find another reference firm with less slack. This may cast some doubt on the whole analysis and make it difficult to argue that the analysis takes into account all possible reference units in determining which will have the largest positive effect on the firm in question. However, the two-stage approach has other drawbacks. Because it maximizes slack, one might argue that it identifies a dominating, fully efficient reference unit, the one farthest away from our firm. There are other ways to select more similar reference firms, though they are more cumbersome.

Another problem with the two-stage approach is that it varies based on the units of measurement used. If we measure an input in tons rather than kilograms, for example, our choice will affect the resulting measures and, more importantly, the

choice of reference firm that we might present to a client. One way to overcome this problem is to look at slack in relative rather than absolute terms. We do this in the following problem, where we measure slack relative to the corresponding input or output:

$$\begin{aligned}
\min_{E,\lambda,z^-,z^+} \quad & E - \frac{1}{m}\sum_{h=1}^{m}\frac{z_h^-}{x_h^o} - \frac{1}{n}\sum_{i=1}^{n}\frac{z_i^+}{y_i^o} \\
\text{s.t.} \quad & \sum_{k=1}^{K}\lambda^k x_h^k = E x_h^o + z_h^-, && h = 1,\ldots,m \\
& \sum_{k=1}^{K}\lambda^k y_i^k = y_i^o - z_i^+, && i = 1,\ldots,n \\
& \lambda \in \Lambda^K(\gamma), \quad z^- \geq 0, \quad z^+ \geq 0
\end{aligned}$$

Again, we can use this principle in a two-stage approach, i.e. we first minimize E and then maximize the sum of the average, relative slack values. As a result, the aggregated slack value does not depend on the measurement scale. Of course, one can still argue that the weighting of the slack is somewhat arbitrary.

Before closing this discussion of slack with a numerical example, we note that the issue of slack perhaps should be less concerned with the measurement task, and more with the modeling task. Recent research suggests that slack may not just be a function of the piecewise linear approximations in DEA. Excessive slack may also indicate that the model specification is wrong and that what we are modeling as joint production is in fact a combination of *independent sub-processes* in which some inputs are used to produce particular outputs while other inputs are used to produce other outputs. In such cases, genuine inefficiency in one process will appear to be slack in the combined process even in cases in which we do not have vertical or horizontal frontier segments in the underlying technologies. This research is still not very well developed, but our results seem to see considerable amounts of slack as a warning that our view of the production process may not be accurate.

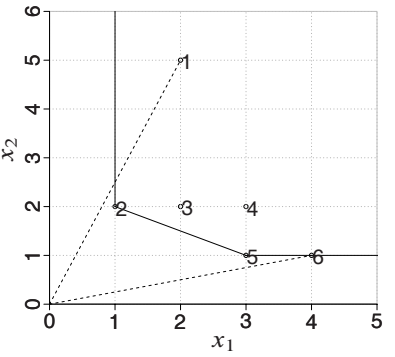
Numerical example in R

Consider an example with six firms that have used two inputs to produce one output. The inputs are given in Table 5.4, and we assume that they have all produced the same output, which we will designate as y=1. If we use a VRS DEA model, we obtain the Farrell input efficiencies shown in the E column. In addition, if we use the slack-based modified efficiency score, we get the efficiency scores shown in the E^* columns when we use $\delta = 0.1$.

Now, as noted above, the value of E^* will depend on the size of δ, and it is generally better just to report E and possibly the individual slacks. We can do this

Table 5.4 Efficency is penalized by Slack in input

Firm	x_1	x_2	E	$E^*(\delta = 0.1)$
1	2	5	.50	.45
2	1	2	1.00	1.00
3	2	2	.83	.83
4	3	2	.71	.71
5	3	1	1.00	1.00
6	4	1	1.00	.90



using the slack option in the `(dea)` function in benchmarking or, alternatively, the `slack` function. The following R code illustrates the first approach.

```
> x <- matrix(c(2,1,2,3,3,4,5,2,2,2,1,1), ncol=2)
> y <- matrix(c(1,1,1,1,1,1), ncol=1)
> e_vrs <- dea(x,y,RTS="vrs",ORIENTATION="in",SLACK=TRUE)
> e_vrs$eff
[1] 0.5000 1.0000 0.8333 0.7143 1.0000 1.0000
> e_vrs$slack
[1]  TRUE FALSE FALSE FALSE FALSE  TRUE
> e_vrs$sx
     sx1 sx2
[1,]   0 0.5
[2,]   0 0.0
[3,]   0 0.0
[4,]   0 0.0
[5,]   0 0.0
[6,]   1 0.0
> e_vrs$sy
     sy1
[1,]   0
[2,]   0
[3,]   0
[4,]   0
[5,]   0
[6,]   0
```

We see that firms 1 and 6 have slack, whereas the other firms do not. These results correspond to the graphical illustration; firms 1 and 6 are projected on vertical and horizontal segments. We can derive the exact slack values on the input side from the `sx` portion of the `dea` output. We see that firm 1 has half a unit of slack in the second input, whereas firm 6 has one unit of slack in the first input. We also see that there is no slack on the output side in this example.

5.7 Measurement units, values and missing prices

DEA-based efficiency estimates do not depend on the measurement scale that we use for the different inputs or outputs. It does not matter if we measure a variable in kg or in tons or if we measure another variable in consumers or millions of consumers. Essentially, the results are invariant to positive linear transformations; i.e. we can transform any input x_i with a positive number $\alpha_i > 0$

$$x_i \to \alpha_i x_i, \qquad i = 1, \ldots, m$$

and any output by a positive number $\beta_j > 0$

$$y_j \to \beta_j y_j, \qquad j = 1, \ldots, n$$

without affecting the efficiency measure, the peers or any other essential information extracted from the DEA programs. The set of feasible values $(e, \lambda_1, \ldots, \lambda_K)$ is not affected by such transformation; i.e. the set of feasible solutions to

$$\begin{aligned} \min_{E,\lambda^1,\ldots,\lambda^K} \; & E \\ \text{s.t.} \quad & E\alpha_i x_i^o \geq \sum_{k=1}^{K} \lambda^k \alpha_i x_i^k, \quad i = 1, \ldots, m \\ & \beta_j y_j^o \leq \sum_{k=1}^{K} \lambda^k \beta_j y_j^k, \quad j = 1, \ldots, n \\ & \lambda \in \Lambda^K(\gamma) \end{aligned}$$

is the same as to

$$\begin{aligned} \min_{E,\lambda^1,\ldots,\lambda^K} \; & E \\ \text{s.t.} \quad & E x_i^o \geq \sum_{k=1}^{K} \lambda^k x_i^k, \quad i = 1, \ldots, m \\ & y_j^o \leq \sum_{k=1}^{K} \lambda^k y_j^k, \quad j = 1, \ldots, n \\ & \lambda \in \Lambda^K(\gamma) \end{aligned}$$

because we can simply divide through by α_i and β_j for the respective constraints. Of course, for this to work, we must use the same transformation of a given variable across all firms; i.e. we cannot measure production in tons for some firms and in kg for others.

It is worth noting that it is only such linear transformations that have no impact. We cannot generally add a constant without affecting the outcome. Moving from x_i

to $a + bx_i$, for example, will affect the outcome if a is not equal to 0. In such cases, we must use particular efficiency measures (e.g. directional distances) to obtain invariance.

One application of this observation is to the use of monetary values as opposed to physical values. When data come from *accounting data*, they are mostly values. and therefore. quantities and prices will be missing. That is, instead of physical inputs $(x_1^k, \ldots, x_n^k)$ and outputs $(y_1^k, \ldots, y_m^k)$ and corresponding input prices $(w_1, \ldots, w_n)$ and output prices $(p_1, \ldots, p_m)$, we have only values for inputs $(w_1 x_1^k, \ldots, w_n x_n^k)$ and values for outputs $(p_1 y_1^k, \ldots, p_m y_m^k)$. Note that the way we have written the values, we assume that the firms have the same prices. Now, the results above show that we get the same results using $(w_1 x_1^k, \ldots, w_n x_n^k)$ and $(p_1 y_1^k, \ldots, p_m y_m^k)$ as inputs and outputs as we would if we used $(x_1^k, \ldots, x_n^k)$ and $(y_1^k, \ldots, y_m^k)$ as inputs and outputs.

It follows that we can in fact measure technical efficiency even if we just have *monetary values* (e.g., *cost* values or shares and *revenue* values or shares) as long as we can still distinguish between the cost values or shares of the different input categories and the revenue values or shares from different outputs.

It is also clear that if we are working with cost shares and are looking to evaluate allocative efficiency, then we should set all prices to 1 because they now give the relative costs of the different cost shares and they are all the same because they are already in monetary terms. Similarly, if we have revenue shares and want to evaluate output allocative efficiency, we should also set prices to 1 because the "value" of $p_j y_j^k$ is $p_j y_j^k$. Hence, if we want to calculate allocative efficiency, we should either use the prices of the original inputs or outputs or use the unit prices for the cost shares or revenue shares.

5.8 Dual programs

The mathematical programs that we use to evaluate DEA efficiencies can be reformulated using duality theory. This method can provide us with alternative interpretations that some users prefer and a useful perspective on the interaction between the evaluator (i.e., the regulator or authority) and those being evaluated (the firms). We will introduce the basics of such dualization here. We focus on the traditional model, i.e. the VRS, DRS, IRS and CRS DEA models, because duality theory is more complicated in the context of mixed integer problems. An introduction to linear programming duality is given in Appendix 4.11.2.

Recall that input-based Farrell efficiency in the DEA model E can be calculated as the solution to the LP problem

$$\begin{aligned}\min_{E,\lambda^1,\ldots,\lambda^K} \; & E \\ \text{s.t.} \quad & Ex^{\mathrm{o}} \geq \sum_{k=1}^{K} \lambda^k x^k, \\ & y^{\mathrm{o}} \leq \sum_{k=1}^{K} \lambda^k y^k, \\ & \lambda \in \Lambda^K(\gamma).\end{aligned}$$

which we may rewrite as

$$\begin{aligned}\min_{E,\lambda^1,\ldots,\lambda^K} \; & E \\ \text{s.t.} \quad & Ex^{\mathrm{o}} - \sum_{k=1}^{K} \lambda^k x^k \geq 0 \\ & \sum_{k=1}^{K} \lambda^k y^k \geq y^{\mathrm{o}}, \\ & \lambda \in \Lambda^K(\gamma).\end{aligned} \tag{5.2}$$

In the VRS, DRS, IRS and CRS cases, this is a simple LP problem, and we can therefore determine the dual LP problem. Moreover, the primal problems have a finite solution, and therefore, so does the dual problem, and they coincide; cf. the duality theorem of LP in Appendix 4.11.2.

To find the dual program, recall that the dual of an LP problem of the form $\max_x\{cx \mid Ax \leq b, x \geq 0\}$ is $\min_y\{by \mid yA \geq c, y \geq 0\}$, where c and x are n dimensional vectors, b and y are m vectors and A is a $m \times n$ matrix. Note that we have used the most common symbols in LP theory, x, y, c, b, A, n, and m, as generic vectors, matrices and numbers without any relationship to their meaning in DEA. In addition, to help formulate the dual, some may find it useful to write the above problem of finding E in the VRS case in the LP tabula as

	E	λ_1	$\cdots$	λ_K		max
u	x^{o}	$-x^1$	$\cdots$	$-x^K$	$\geq$	0
v	0	y^1	$\cdots$	y^K	$\geq$	y^{o}
ϕ	0	1	$\cdots$	1	=	1
min	1	0	$\cdots$	0		

Either way, the dual problem can be written as

$$\begin{aligned} \max_{u,v,\phi} \quad & vy^{\mathrm{o}} + \phi \\ \text{s.t.} \quad & ux^{\mathrm{o}} \leq 1 \\ & -ux^k + vy^k + \phi \leq 0, \qquad k = 1, \ldots, K \\ & \phi \in \Phi(\gamma) \end{aligned} \tag{5.3}$$

where $\Phi(\text{vrs}) = \mathbb{R}, \Phi(\text{drs}) = \mathbb{R}_-, \Phi(\text{irs}) = \mathbb{R}_+$, and $\Phi(\text{crs}) = \{0\}$. For the CRS technology in the DEA problem in Eq. (5.2), there are no restrictions on $(\lambda^1, \ldots, \lambda^K)$, and therefore, $\phi = 0$ in the dual problem in Eq. (5.3) and the problem become

$$\begin{aligned} \max_{u,v,\phi} \quad & vy^{\mathrm{o}} \\ \text{s.t.} \quad & ux^{\mathrm{o}} \leq 1 \\ & -ux^k + vy^k \leq 0, \qquad k = 1, \ldots, K \\ & \phi \in \Phi(\gamma). \end{aligned}$$

These problems correspond to a classical *pricing problem*. We choose shadow prices u and v for the inputs and outputs to make the output value of observations (x, y) look as good as possible, create as large a value of vy^{o} as possible, and stipulate that no firm can have generated a net profit $vy^k - ux^k \leq 0$. The $ux^k \leq 1$ condition is simply a norming condition that prohibits infinite solutions that would otherwise be possible by inflating the v values, i.e. making v equal to infinity.

The scalar ϕ is the cost for only having access to convex combinations and not having constant returns to scale. For, given a input x, the higher the value of ϕ, the lower the value of vy necessary to fulfill the restrictions.

We note that what we call the dual problem here is often called the *multiplier model*, while what we have called the primal problem is called the *envelopment model*. Moreover, we note that if we dualize the slack model in Eq. (5.1), then we get a similar DEA except that the multipliers are restricted so that they are no less than δ; i.e., in the multiplier version, $u_i \geq \delta, i = 1, \ldots, m, v_j \geq \delta, j = 1, \ldots, n$.

A *geometrical interpretation* of the dual programs is easiest if we think of a cost function with input x on the vertical axis and outputs y on the horizontal axis. The dual program estimates alternative affine cost functions where the realized costs are higher for all firms than the estimated costs; i.e., $ux^k \geq \phi + vy^k$ for all firms. That is, we are looking at all possible affine functions that are below the observations and attempting to pick the function that maximizes the cost level $vy^{\mathrm{o}} + \phi$ that we can assign to firm o. This is illustrated in the left panel of Fig. 4.5.

Let us return to the dual problem of Eq. (5.3). It is clear that we may choose $ux^{\mathrm{o}} = 1$ without loss of generality and thus rewrite the program as the following non-linear *ratio problem*:

$$\begin{aligned} \max_{u,v,\phi} \quad & \frac{vy^{\mathrm{o}} + \phi}{ux^{\mathrm{o}}} \\ \text{s.t.} \quad & \frac{vy^{k} + \phi}{ux^{k}} \leq 1, \qquad k = 1, \ldots, K \\ & \phi \in \Phi(\gamma) \end{aligned} \tag{5.4}$$

In the CRS case with $\phi = 0$, this is, in fact, the original DEA problem suggested in the seminal papers introducing DEA, as we will return to below. The idea is that we seek to choose prices or priorities so as to aggregate the benefits (output) and costs (inputs). We choose the priorities (prices, value) u and v to maximize the evaluated firm's benefit–cost ratio subject to the condition that no unit can have a higher ratio than 1 with the selected weights. Technically, this program is not an LP program but is rather a fractional linear program.

Although the primal formulation, the envelopment problem, is now used more widely in the literature, the original formulation does have merits that benefit economists, engineers and laymen alike.

From an economic perspective, we see that the evaluation problem is like a *cost-benefit analysis* in which we seek to maximize the benefit-cost ratio. The challenge of this approach is often to determine appropriate weights or values to assign to the otherwise incompatible input and output dimensions. The DEA problem addresses this issue by generating its own endogenous prices. Moreover, the value weights ϕ, u and v selected by the DEA program put the evaluated unit in the best possible light compared to the other units. The dualization thus supports the popular view that *DEA puts everyone in the best possible light*. We have already identified another formal result that supports this perspective, the use of minimal extrapolation technologies, as we have discussed at some length in Chap. 4.

From an *engineering perspective*, efficiency is usually a question of ensuring that high outputs accrue from low inputs. Many measures developed by engineers and economists in different sectors involve such ratios of outputs to inputs (see our discussion of key performance indicators in Chap. 1). In our framework, this concept does not work directly due the multiplicity of inputs and outputs; the ratio will depend on which inputs and outputs are compared. However, we see that the DEA model overcomes this problem by finding its own weights and making the (compound) output to (compound) input ratio look as good as possible.

So far, we have focused on dualizations of the input contraction problem. However, we note that similar dualizations of the Farrell output measure are also possible. Thus, if we consider the Farrell output efficiency problem:

$$\begin{aligned}
\max_{F,\lambda^1,\ldots,\lambda^K} \quad & F \\
\text{s.t.} \quad & x^{\mathrm{o}} \geq \sum_{j=1}^{K} \lambda^k x^k \\
& F y^{\mathrm{o}} \leq \sum_{j=1}^{K} \lambda^k y^k \\
& \lambda \in \Lambda^k(\gamma).
\end{aligned}$$

we see that the dual, the multiplier version, becomes

$$\begin{aligned}
\min_{u,v,\psi} \quad & u x^{\mathrm{o}} + \psi \\
\text{s.t.} \quad & v y^{\mathrm{o}} \geq 1 \\
& u x^k - v y^k + \psi \geq 0, \qquad k = 1, \ldots, K \\
& \psi \in \Psi(\gamma)
\end{aligned} \tag{5.5}$$

where $\Psi(\text{vrs}) = \mathbb{R}$, $\Psi(\text{drs}) = \mathbb{R}_+$, $\Psi(\text{irs}) = \mathbb{R}_-$, and $\Psi(\text{crs}) = \{0\}$.

Here we find an affine approximation of the production function that makes the least optimistic prediction regarding the feasible output when x^{o} is used and such that all observed production plans are still feasible. This corresponds to the illustration in the panel on the right side of Fig. 4.5.

Again, we can rewrite the multiplier version into ratio form by noting that we may restrict output prices without loss of generality such that $v y^{\mathrm{o}} = 1$; we get

$$\begin{aligned}
\min_{u,v,\psi} \quad & \frac{u x^{\mathrm{o}} + \psi}{v y^{\mathrm{o}}} \\
\text{s.t.} \quad & \frac{u x^k + \psi}{v y^k} \geq 1, \qquad k = 1, \ldots, K \\
& \psi \in \Psi(\gamma)
\end{aligned}$$

In this book, we have introduced DEA based on production theory. We assume some underlying technology T that we try to estimate based on the data and a few basic assumptions from production economics. Indeed, we would consider this the natural and standard mode of operation at this point. However, it is interesting to note that DEA was originally developed using a ratio formulation intended to solve a weighting problem. As such, it does not directly assume any underlying technological properties but rather simply compares existing units. The way in which we construct the ratios, using only linear weightings, indicates that this has a natural dual formulation as an activity analysis problem that is well-known from production theory. When Charnes, Cooper and Rhodes first presented their methods, they used the fractional programming ratio problem as the primary formulation and linearized it. If the linearized problem is dualized, we obtain a problem equivalent to

an activity analysis, the envelopment form, which is the problem we now consider to be the primal one.

5.9 Maximin formulations

The DEA programs can also be reframed as maximin programs. This approach is interesting because it may yield new interpretations and shed light on the nature of the DEA measures and their possible use in strategic contexts.

In the ratio form of the Farrell input efficiency problem Eq. (5.4), it is clear that the highest output-input ratio will also be 1. Otherwise, we can improve the objective via a small proportional expansion of all the weights in (ϕ, v). We can therefore also reformulate the program as the following equivalent *maximin program*

$$E^k = \max_{u,v,\phi} \min_k \frac{\frac{\phi + vy^o}{ux^o}}{\frac{\phi + vy^k}{ux^k}}$$

This suggests that we can look at the scoring problem as a *game problem* in which the evaluated and the evaluator are the participants. The *firm being evaluated chooses the priorities (u, v and ϕ) based what it wishes to be evaluated, and the evaluator selects a comparator* ($k \in \{1, \ldots, K\}$). The firm being evaluated seeks to make his benefit-cost ratio appear as high as possible, whereas the evaluator seeks to make it look the least impressive by identifying better practices given the priorities selected. In practice, the firm being evaluated does not really specify the priorities—the DEA program does so endogenously—but this is only to the advantage of the former because it could not have chosen the priorities in any better way.

If we assume that the technology is the DEA CRS model and that all units have produced the same outputs y, the maximin program simplifies because $\phi = 0$ and the output factors cancel out, such that

$$E^o = \max_u \min_k \frac{ux^k}{ux^o}$$

The focus of this program is the cost of the input x^k in an alternative plan against the cost of the input x^o for the evaluated firm. The firm prefers this cost to be high, whereas the evaluator prefers it to be low. We therefore see that the evaluation is like a game in which the former seeks to maximize the cost ratio by picking appropriate prices or priorities and the latter seeks to minimize it by choosing the appropriate comparator $k \in \{1, \ldots, K\}$. Because the two parties have directly opposing interests, we can think of this as a *zero-sum game*.

5.10 Partial value information

In the efficiency analyses we have discussed so far, we have either assumed that no price information is available and focused on technical efficiency, or we have assumed that exact prices are available, thereby measuring cost efficiency, revenue efficiency or profit efficiency.

In some situations, however, we have partial value or price information; i.e., we have some price information, but our information is imperfect. We will now discuss how to incorporate such information into the efficiency programs.

Assume that we have some prior view regarding the relative worth of inputs or and outputs. In a hospital setting, for example, we may know that the value of one heart operation exceeds that of one knee operation, but we may not know the more precise relative worth of the two treatments. Still, we may be able to use such partial information to refine the efficiency evaluations. Imagine a situation in which we only have two hospitals. Hospital A has conducted 100 knee operations, and hospital B has conducted 150 heart operations. Both have the same total costs. Now, in a single-input, two-output model, the two hospitals would both be considered to be fully Farrell efficient. Introducing our partial value information, however, we can say that the output value of hospital B exceeds the output value of hospital A by at least 50%, and we can therefore say that the output efficiency of hospital A can be at most $\frac{150}{100} = 1.5$.

The DEA literature makes several suggestions regarding how to include partial value information in such evaluations. We will mainly focus on the most popular approach, the use of *assurance regions*. The idea here is to introduce *weight restrictions* as part of the dual formulations, i.e. restrictions on dual prices.

The simplest way to do so, sometimes referred to as the creation of Type 1 assurance regions, is to restrict relative input prices or relative output prices using simple restrictions such as

$$\alpha_{h,i} \leq \frac{u_h}{u_i} \leq \beta_{h,i}$$
$$\alpha^*_{h,j} \leq \frac{v_h}{v_j} \leq \beta^*_{h,j}$$

The first restriction indicates that the relative worth of input h to input i is at least $\alpha_{h,i}$ and at most $\beta_{h,i}$. The second restriction is similar on the output side; the relative worth of output h to output j is at least $\alpha^*_{h,j}$ and at most $\beta^*_{h,j}$. In a hospital setting, for example, we could say that

$$0.5 \leq \frac{u_{\text{physician}}}{u_{\text{nurse}}} \leq 4$$
$$1 \leq \frac{v_{\text{heart}}}{v_{\text{knee}}} \leq 10$$

That is, the cost of a physician relative to a nurse is at least 0.5 and at most 4, and the value of a heart surgery is at least the same as that of a knee surgery and at most equivalent to 10 knee surgeries.

We can also use more advanced versions:

$$\sum_{i=1}^{m} \kappa_i u_i \leq 0$$
$$\sum_{j=1}^{n} \kappa_j^* v_j \leq 0$$

where we restrict the relative worth of more than two inputs or outputs at the same time or create restrictions, sometimes called Type II assurance regions, that include both inputs and outputs

$$\hat{\alpha}_{i,j} \leq \frac{u_i}{v_j} \leq \hat{\beta}_{i,j}$$

In general, these more complicated restrictions are more difficult to interpret and to justify.

Assurance regions like the above can generally be interpreted in two different ways.

One is as *expressing preferences* in terms of subjective values assigned to inputs and outputs. We can also say that the partial values allow us to partially transform the technical efficiency evaluation into a utility-based effectiveness analysis. This is perhaps most clear from the maximin formulations. The restrictions on dual weights restrict the objectives that the evaluated firm can possibly claim. In the example, a hospital cannot reasonably claim that one knee operation is more valuable than one heart operation, which restricts hospital A's ability to make itself appear efficient.

This view on dual weight restrictions is also related to the literature linking DEA and *MCDM, Multiple Criteria Decision-Making*. One relevant approach is value efficiency analyses (VEA), in which a decision-makers preferences are partially revealed through his preferred production plan. The preference function serves the same purpose as market prices because it allows us to aggregate inputs and outputs.

Another interpretation of the dual restrictions is as an *expression of technical rates of substitution*. Consider the dual version of the Farrell input efficiency program, Eq. (5.3). We have

$$-ux^k + vy^k + \phi \leq 0, \qquad k = 1, \ldots, K$$

In optimum, at least one of these will be binding for firm o; at least one of the λ values will be positive. A facet or hyperplane will emerge that firm o is projected against, namely, the set of (x, y) values for which $-ux + vy = k^1$ where k^1 is a constant. Fixing the values of x, we see that the hyperplane in the output space defines the approximate output possibility set given by $vy = k^2$, where k^2 is another constant. We can therefore determine the rate of technical transformation between y^h and y^j as

$$\frac{\mathrm{d}y^h}{\mathrm{d}y^j} = -\frac{v^j}{v^h}$$

Hence, if we want to increase y^j by 1 unit, we will have to reduce y^h by $\frac{v^j}{v^h}$. We therefore see that assurance regions can be interpreted as restricting the rate of technical transformation on the output side. Likewise, if we look at the input side, we get $ux = k^3$ such that

$$\frac{\mathrm{d}x^h}{\mathrm{d}x^i} = -\frac{u^i}{u^h}$$

which suggests that assurance regions on the input side restrict the rate of technical substitution between the production factors. If we use one more of input i, we can save $\frac{u^i}{u^h}$ of input h. It follows that we can also consider partial restrictions on the dual weights as extensions of the production possibilities.

In fact, one can take this last idea a step further and consider the introduction of artificial observations into the primal space or the transformation of the primal inputs and outputs before an efficiency analysis is undertaken. One set of results along these lines is developed for cone ratio extensions of the CRS model.

Numerical example in R

Consider a situation in which four medical teams using nurses and physicians conduct knee and heart surgery. The inputs and outputs of the four teams are given in Table 5.5.

Table 5.5 Medical teams

Team	Nurses	Physicians	Knees	Hearts
A	3	3	100	0
B	3	3	0	150
C	6	2	50	75
D	1	4	50	75

Now, using `dea.dual`, we can calculate the dual weights and we can also add restriction on these as illustrated in the following R code. We use `e` for the result from the ordinary DEA analysis and `edr` for the restricted dual DEA analysis.

```
> library(Benchmarking)
> x <- matrix(c(3,3,6,1,3,3,2,4), ncol=2)
> y <- matrix(c(100,0,50,50,0,150,75,75),ncol=2)
> e <- dea(x,y,RTS="crs")
> cbind(E=e$eff, e$ux, e$vy)
     E         u1         u2   v1          v2
[1,] 1 0.27777778 0.05555556 0.01 0.000000000
[2,] 1 0.11111111 0.22222222 0.01 0.006666667
[3,] 1 0.08333333 0.25000000 0.01 0.006666667
[4,] 1 0.11111111 0.22222222 0.01 0.006666667
```

```
> dual <-matrix(c(0.5,1, 4,10),ncol=2)
> dual
     [,1] [,2]
[1,]  0.5    4
[2,]  1.0   10
> edr <- dea.dual(x,y,RTS="crs", DUAL=dual)
> cbind(E=edr$eff, edr$u, edr$v)
             E         u1        u2          v1          v2
[1,] 0.6666667 0.16666667 0.1666667 0.006666667 0.006666667
[2,] 1.0000000 0.16666667 0.1666667 0.006666667 0.006666667
[3,] 0.8928571 0.07142857 0.2857143 0.007142857 0.007142857
[4,] 1.0000000 0.20000000 0.2000000 0.008000000 0.008000000
```

We see that without restrictions on the dual variables, they all emerge as efficient. From the dual values, we can also see that this is explained in part by the zero value that team D assigns to heart surgeries; this is clearly not realistic. In addition, we can see that other medical team assigns a relatively lower value to nurses compared to physicians than does team D.

Now, we can restrict the input and output values using the assurance regions suggested above. We do this using the matrix `dual`, which indicates the input prices in relation to the first input with the lower bound in the left column and the upper bound in the right column. The assurance region for the output prices is similar. On this basis, we see that team A is no longer efficient. The reason is that A must assign at least the same weight to heart surgery that it does to knee surgery. In doing so, team B has been able to produce 50% larger outputs with the same inputs. In CRS, input efficiency is the inverse of output efficiency, and therefore, the efficiency of team A becomes 100/150=0.667. Team C is also no longer efficient. It chooses to make nurses four times more expensive than doctors. Thus, it cannot be dominated by team D, but it can be dominated by team B. Using a weight of 1 for doctors, a weight of 4 for nurses and equal weights for knee and heart surgery, teams B and C have used inputs of $3 \cdot 1 + 3 \cdot 4 = 15$ and $6 \cdot 1 + 2 \cdot 4 = 14$ to produce output values of $0 \cdot 1 + 150 \cdot 1 = 150$ and $50 \cdot 1 + 75 \cdot 1 = 125$. The input efficiency of team C as compared to team D is therefore $(15/150)/(14/125) = 0.893$. Note that in the calculations, we have used the dual variables for team C in accordance with the interpretations above. The dual problem (with or without restrictions) is used to find the values of the inputs and outputs that put the team in question in the best possible light.

Note also that for ease of explanation, we did not use the dual values directly but instead rescaled them. Thus, for example, we said that team C uses an input weight ratio of $1 : 4$ instead of $0.07142857 : 0.2857143$. Such changes do not affect the DEA programs because it does not matter which units we use for the different inputs and outputs.

To avoid any confusion, however, let us also do the calculations using the dual weights directly. Based on the dual weights for team C, the four teams have produced the aggregated inputs and outputs shown in Table 5.6.

If we do the efficiency calculations in a one-input, one-output model, team C emerges as having an efficiency level of 1 before the weight restrictions and 0.89

Table 5.6 Evaluation of Team C

Team	Unrestricted		Restricted	
	Inputs	Outputs	Inputs	Outputs
A	1.00	1.00	1.07	0.71
B	1.00	1.00	1.07	1.07
C	1.00	1.00	1.00	0.89
D	1.08	1.00	1.21	0.89

= (1.07/1.07) /(1.00/0.89) after the restrictions. The efficiency scores of the other teams can be explained in a similar way. We just have to use the specific dual weights for each team, reflecting the idea that DEA chooses weights for the individual firms that make them look as positive as possible.

5.10.1 Establishing relevant value restrictions

The challenge in applying this procedure is of course to establish restrictions on the input and output weights that makes sense or to suggest hypothetical production plans that can generally be accepted.

One approach is to use information on prices or costs. The relative worth of outputs may in some cases be estimated using existing market prices or market prices for related services. Because prices often vary over time and based on location, and because specific resources and services may not be priced individually, if is often more realistic to extract price ratio intervals as used in the assurance region approach than to extract relative prices as used in costs and revenue efficiency analysis.

Another approach is to use expert opinions. Again, these will typically vary, and instead of averaging them, it is often safer to create a consensus based on some interval estimates.

A third approach is to use models and methods from accounting, engineering or statistics to determine possible aggregations of different services or resources. Because such models are typically somewhat uncertain, the extracted information may best be used as partial information.

Practical application: Regulation

To refine regulatory benchmarking models, a series of supplementary approaches are typically used to at least partially establish input or output values that can be used to aggregate the inputs and outputs. This step helps to combat the pressure to include many details in models estimated based on relatively few data points.

We have already discussed the introduction of quasi-observations (i.e., hypothetical production plans) in Sect. 4.7. Such observations can be derived from engineer-

ing models, accounting analyses or management consultants in-depth analyses of actual firms opportunities for improvement.

Another more common strategy is to estimate *unit costs or cost equivalents*. In a network, the main cost drivers are typically the different assets (e.g., the km of lines of different voltage classes and the different types of transformers). It is impossible to estimate their contribution to costs directly using DEA or econometric techniques because the number of observations is typically small (e.g., 20-200), whereas the number of different asset types is large, (e.g., 20-1500). Instead, relative costs are estimated using either cost allocation rules from accounting or engineering models calibrated to projects where detailed cost information is available. Once the relative weight, the unit costs or the cost equivalents are established, we can construct a few cost-aggregated "size of grid" or "netvolume" measures, such as

$$\text{Netvolume}(g) = \sum_{k=1} K(g) v_k N_k, \qquad g = 1, \ldots, G$$

where $k = 1, \ldots, K(g)$ are the different assets in group g (say, lines), N_k is the number of assets of type k, and v_k is the relative costs of these assets compared to that of other assets in the same group. In the DEA model, one can then use the Netvolumes(g), $g = 1, \ldots, G$, as the main cost drivers. This means that we restrict the relative prices inside the groups but let the DEA model determine the relative weighting of the different groups.

A specific example involving this approach is the *e3GRID* benchmarking project that was conducted for 22 national transmission system operators (TSO) from 19 different countries commissioned by Council of European Energy Regulators (CEER) on behalf of the national regulatory authorities. The overall objective of the project was to deliver static and dynamic cost efficiency estimates that would be robust and understandable and could be used with a variety of regulatory applications, from comprehensive performance assessments to structured periodic rate reviews (e.g., in setting X-factors). The efficiency estimation techniques used depended on the character of the underlying functions in terms of homogeneity, cost causality and production space. The most extensive assessment was made using a non-parametric DEA frontier model under the assumption of non-decreasing returns to scale and encompassing total expenditure for construction, maintenance, planning and administration (CMPA). More than 1,200 different assets were identified by the TSO, and therefore, extensive aggregation was necessary with only 22 observations from each of the 3 years. In addition to measuring grid volume, the model also included density and decentralized generation capacity in the network as cost drivers.

5.10.2 Applications of value restrictions

The inclusion of partial value information has several applications. Let us mention just some of them.

One advantage (and perhaps the main advantage) is that including this information allows efficiency analyses to contribute to *effectiveness or value for money analyses*; i.e., it allows us to make rational ideal evaluations. This presumes, of course, that the restrictions reflect the perceived relative worth of inputs or outputs.

One might also suggest that obvious information about substitution rates for inputs or outputs should be included because this allows a more fair and correct evaluation. In this case, we need partial information reflecting technological possibilities rather than values.

Another advantage is that including this information allows us to *work with more inputs and outputs* even in cases with a limited number of observations. Studies have shown that if primal information about inputs and outputs is supplemented with just partial dual information about the relative importance of different types of inputs and outputs, the number of inputs and outputs can be expanded considerably.

Another advantage might be the reduction of *bias in the efficiency estimates*. We will discuss bias in more detail in the next chapter. The basic idea, however, is simple: the minimal extrapolation principle includes a bias because we develop an inner approximation of the underlying true production possibility set. This means that the true efficiencies are lower than the relative efficiencies we estimate. This bias is particularly large in those parts of the production space where we have relatively few observations, and one might expect partial price information, such as more elaborate rescaling possibilities, to reduce this problem.

Finally, such information can help us to evaluate what otherwise appear to be hyper-efficient firms. By supplementing VRS DEA models with more general return NDRS or CRS models or by including weight restrictions, we can eliminate some of these problems.

In applications, the main disadvantage of the use of value restrictions is that *interpretations become less clear*. The implicit targets that correspond to the projections, i.e.

$$(E^{*o}x^{o}, y^{o}) \text{ or } (x^{o}, F^{*o}y^{o})$$

where E^{*o} and F^{*o} are the Farrell input and output efficiencies calculated in models with partial value information, may lie outside the production possibility set spanned by the original observations. There is no direct empirical evidence that they are feasible, and there is no simple combination of best practices on which to rely. Needless to say, this eliminates one of the most compelling merits of DEA compared to other approaches: its reliance on minimal extrapolation from best practices and its ability to point to a few peer firms.

We may also consider weight restrictions and dual problems with partial price information in terms of *sensitivity analysis*. Under partial value information, we are uncertain about the appropriate weighting of the inputs and outputs but we do know something. In serious applied studies, it is common to investigate how sensitive the results are to the main uncertainties. In practice, this may be difficult because we need to allow for simultaneous uncertainty about many parameters. The dual program with such value restrictions is an advanced way to investigate the sensitivity of the results. It is basically used to investigate all possible remaining combinations

of prices so as to determine exactly the combination that puts each firm in its best possible light.

5.11 Summary

In this chapter, we have introduced the idea of super-efficiency as firm efficiency relative to the technology spanned by other firms. A super-efficient firm can increase its inputs or reduce its outputs at least somewhat without appearing inefficient in a traditional efficiency analysis. This concept helps us to discriminate among efficient firms and can help to guide regulations that require incentives for even the most efficient firms.

We have also introduced sub-vector efficiency as a way of addressing scenarios in which only some of the inputs or outputs are discretionary. We have shown how easily this process is implemented in the DEA models. Likewise, we have shown how the idea of directional distance functions, where improvements are sought in arbitrary directions, can be directly implemented in a DEA context using linear and mixed integer programming.

We have briefly discussed possible slack in individual inputs and outputs for firms that are Farrell efficient, and we have covered some possible modifications to the efficiency measure. The slack problem is quite common in DEA models due to the vertical and horizontal segments of the frontier. We have argued that it normally works well to simply report the Farrell efficiencies and the possible slack determined based on a second-stage maximization of the slack after radial improvements.

We have observed also that DEA problems, at least if we do not included slacks adjustments, are invariant to positive linear transformations of the inputs and outputs. That is, the unit of measurement of any given variable does not matter. This also means that when firms face similar prices, we can make the technical efficiency analyses using values such as cost and revenue shares.

Finally, we have discussed the dual version (the multiplier form) of the usual Farrell efficiency programs in the VRS, DRS, IRS and CRS cases. The dual problems have nice economic interpretations as pricing problems. They can also be rewritten in ratio form and thus provide an alternative interpretation of DEA as a cost-benefit analysis in which we lack ex ante priorities, prices or values to aggregate the costs and benefits. Instead, we choose the priorities endogenously in the evaluation process to make the evaluated firm look as good as possible. Lastly, the dual problem can be rewritten as a minimax problem emphasizing the opposing interests of the evaluated firm and the evaluator in a game-like scenario. The firm being evaluated can be thought of as selecting the relative weights of the inputs and outputs so as to appear as effective as possible. However, given these weights, the evaluator may find alternative firms that are succeeding even better with the same priorities.

We have also shown how the implicit prices or values in a DEA analysis can be extracted from the dual solution and how we can restrain the relative importance of inputs and outputs by restricting the dual variables using assurance regions, for

example. This may allow us to make more relevant evaluations and to better consider best practices.

5.12 Bibliographic notes

Super-efficiency was first suggested by Andersen and Petersen (1993) as a means of differentiating among frontier units. Its role in contracting was first established in Bogetoft (1994a, 1995).

The use of discretionary and non-discretionary dimensions dates back to Banker and Morey (1986)

As mentioned in Chap. 2, early and basic work on the excess function was presented by Luenberger (1992) and Chambers et al (1998).

Potential improvements or multi-dimensional efficiency evaluations were introduced by Bogetoft and Hougaard (1999), extended to super-efficiency in Bogetoft and Hougaard (2004), and applied in Asmild et al (2003), among others. The approach is based on axiomatic bargaining theory along the lines of Kalai and Smorodinsky (1975) 's solution to a two-person bargaining problem.

The linkage between DEA and Multiple Criteria Decision-Making MCDM is discussed in Halme et al (1999) and Joro et al (1998). A flexible approach combining the two literatures is interactive benchmarking suggested in Bogetoft and Nielsen (2005) and Bogetoft et al (2006a).

Dualization of standard DEA problems is discussed in several DEA texbook. Dualizations of non-standard problems, including FDH problems, are discussed in Agrell and Tind (2001) and Kuosmanen (2003). The interpretation of the ex-post scoring problem in DEA as a zero-sum game was first suggested by Banker (1980) and Banker et al (1989). Weight restrictions are discussed in a large number of articles and books. A good survey of assurance regions is Thanassoulis et al (2004). Papers on cone ratio analysis consider the relationship between restricting the dual weight to (polyhedral cones) and making linear transformations of the primal observations before ordinary benchmarking is undertaken. Cone ratio analysis was introduced by Charnes et al (1989). The use of partial price information to vastly expand the the number of inputs and outputs one can include in a DEA model is demonstrated in Olesen and Petersen (2002).

In terms of applications covered in this chapter, the analysis fishery is discussed in Andersen and Bogetoft (2007), the banking example in Asmild et al (2008), relying in part on the rational inefficiency concept by Bogetoft and Hougaard (2003), and the e3GRID project is reported in Agrell and Bogetoft (2009).

Outlier detection, as discussed in the Appendix, is an important topic in applied projects although seldom taught in great details in econometrics classes. General diagnostic methods are treated in Atkinson (1985) and Belsley et al (1980). The use of super-efficiency to find outliers in a DEA context is discussed in Banker and Chang (2006). The details of the data cloud methods can be found in Andrews

and Pregibon (1978) and Wilson (1993). To fully understand the discussions herein, however an advanced knowledge of statistics and distributions is needed.

5.13 Appendix: Outliers

Outliers are firms that differ to a large extent from the rest of firms and therefore may end up being badly captured by the model or having too large an impact on the model. Outliers are helpful when one is using most empirical methods, but they are often thought to be particularly troublesome to DEA because an outlier helps to span the frontier and may have a significant impact on the evaluation of several other firms. We will therefore discuss outliers and ways to identify them in this section.

5.13.1 Types of outliers

There are several reasons a firm may be an outlier.

- First, there may be *errors in the data*. Inputs or outputs may have been lost, there may be errors in typing or punching data values. Such outliers should ideally be corrected or perhaps eliminated because they are do reflect a real production process.
- Secondly, the observations may be potentially correct but highly atypical, sometimes called high leverage points. They may sometimes be identified and eliminated so that the model is not distorted to fit these extreme observations.
- Thirdly, observations that suggest exceptionally low or high relative performance in a parametric or non-parametric model are candidates for outlier detection in benchmarking. In particular, regulatory benchmarking implies that observations that influence the estimations for a large part of the reference set should correspond to replicable firm-level performance for the same set and circumstances. If the relative performance difference is extreme, the individual observation is classified as an outlier in regulatory benchmarking for precautionary reasons, which are not necessarily the same as the second types of outlier detection. On the other hand, such observations could also represent an important phenomenon. They could reflect the first introduction of new technology into a production process or an innovation in management practice from which others would want to learn.

The impact of outliers may also depend on the model. In DEA, particular emphasis is given on the quality of observations used to define best practices. The outlier analysis in DEA can use statistical methods and dual formulation, in which marginal substitution ratios can reveal whether an observation is likely to contain errors. In SFA, outliers may distort the estimation of the curvature. They may also increase

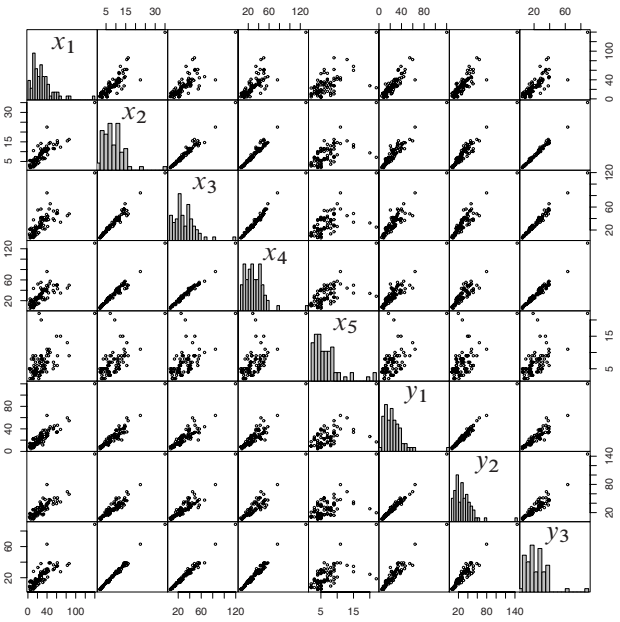


Fig. 5.3 Scatterplot matrix of the Charnes et al (1981) data set `charnes1981.csv`: 5 input and 3 outputs. Histogram of the variables on the diagonal.

the magnitude of the idiosyncratic error term, thus influencing average efficiency estimates in the sample.

5.13.2 Identifying outliers

A simple way to see if there might be a problem with outliers is to make some simple graphs of the data. A useful tool is the *scatterplot matrix*, which in R is available via the function `pairs`.

A simple example with a data set for 70 firms with 5 inputs and 3 outputs from the data set "charnes1981.csv available in the Benchmarking package is shown in figure 5.3. In this graph, there are signs of an outlier: one firm seems to have larger inputs and outputs than the other firms, as on the bottom line, there is a single dot above all the other grouped dots. This simple graphical method is a useful first method of finding outliers, but such graphs are not useful if the extreme features are reflected in a linear combination of more than two variables.

Outliers are also important in regression models, where they can have a large influence on the estimates. Here, we are particular concerned with firms for which a variable is extremely large, meaning that the firm has potential leverage in influenc-

ing the shape and slope of the regression, and that the firm is off-center in the sense that they actually exercise their leverage. There is a whole set of methods that can be used to help identify such outliers. Most of the available methods builds on the residuals from the regressions and are generally designed for normal linear models. They are mostly computational shortcuts used to see how the estimated residual and parameters change when each of the individual observations is excluded one at a time. This is especially true for measures like Cook's distance, DFFITS, partial leverage, conditional numbers, and other methods in which computation depends on the projection or hat matrix. It follows that such methods are not directly relevant to DEA analysis. Some of the methods might be relevant for SFA analysis, but because the SFA estimates also differ from estimates in a normal linear model, they are probably not ideal for SFA models either. Moreover, most of the methods only consider the influence of a single firm, not a group of firms.

We therefore require an approach that is directly focused on identifying outliers that to a large extent may influence frontier models and a method that can preferably handle not only individual outliers, but also groups of outliers.

One approach that has been proposed for DEA models is to look at super-efficiency (cf. subsection 5.2 on page 115) and classify firms as outliers if their input-based super-efficiency is large (say, 3 or 4). The idea of using super-efficiency is straightforward. High super-efficiency means that the firm is significantly pushing out the frontier, and experiments have shown that the method actually works well in practical applications. Unfortunately, its theoretical foundations are limited. Moreover, this method can only find a single-firm outlier and cannot zero in on groups of firms.

Let us therefore turn to a more advanced—and complicated—method.

5.13.3 Data cloud method

Let $X = (x^1, \ldots, x^K)$ and $Y = (y^1, \ldots, y^K)$ be $K \times m$ and $K \times n$ matrices with inputs and outputs for K firms. The combined matrix $\begin{bmatrix} X & Y \end{bmatrix}$ then contains all of our observations. These observations, the different rows in the combined matrix, can be seen as a cloud of points in the $\mathbb{R}^m_+ \times \mathbb{R}^n_+$ space, where each point represents a firm. The volume of the cloud is proportional to the determinant of the combined matrix $\begin{bmatrix} X & Y \end{bmatrix}' \begin{bmatrix} X & Y \end{bmatrix}$:

$$\text{Volume of data cloud } \simeq D(X, Y)\,.$$

It is interesting to note that this determinant can also be interpreted as the generalized sum of the quadratic residuals from the linear model of Y conditioned on X, i.e. the model $\mathrm{EV}(Y|X) = XB$ or

$$Y = XB + \text{noise}$$

where B is a $m \times n$ matrix with parameters.

If we remove a firm from the data, then the volume of the data cloud may decrease. If the removed firm is in the middle of the cloud, the volume will be unchanged. If, on the other hand, the firm is outside the remaining cloud, then the volume will be much smaller, and we will have an indication that the firm is an outlier. To look for one or more outliers, we can therefore look at how the volume of the cloud changes when we remove one or more observations.

Let $D^{(i)}$ be the determinant after removing firm i, and consider the ratio of the new volume of the data cloud to the old volume

$$R^{(i)} = \frac{D^{(i)}}{D}.$$

Note that $R^{(i)}$ does not depend on the units in either the X or the Y matrix; i.e. it is dimensionless. If firm i is not an outlier, then D will not change much and $R^{(i)}$ will be close to 1. If firm i, on the other hand, is an outlier, then $R^{(i)}$ will be much smaller than 1. To look for outliers, we therefore must simply look for small values of $R^{(i)}$. In this approach, we are not restricted to deleting just one observation at a time. We could eliminate firms 1,2 and, 5, for example, and let the resulting ratio of volumes be denoted as $R^{(1,2,5)}$.

To identify outliers or groups of outliers, we must therefore look for small values of R. We do not need all of the small values to make inferences about outliers; we just need to investigate the smallest R for each number of firms that we delete from the data set. Because R is a stochastic variable, we could find its distribution, but to find the distribution of the minimum of R is cumbersome. Let us therefore find another way to look for small values of R.

If there is a group of s outliers and we look for outliers by deleting groups of $1, \ldots, r$ firms, then for $r < s$, we should not expect to find an R with a very small value because there will still be outliers in the remaining data set. However, for $s < r$ we will get an R from which all outliers are deleted, and this R will presumable be very small. When examining the values of R, we will therefore look for the first single isolated small value. If such a value exists, we have found a group of outliers. An isolated small value is an isolated minimum value, or, to fix it on a scale, $\frac{R_{\min}}{R_{\min}} = 1$ should be isolated from other values of $\frac{R^{(r)}}{R_{\min}}$, or 0 should be isolated from other values of $\log\left(\frac{R^{(r)}}{R_{\min}}\right)$. Instead of doing the distributional calculations, we can therefore use a graphical method in which we plot the ordered pairs

$$\left(r, \log\left(\frac{R^{(r)}}{R^{(r)}_{\min}}\right)\right)$$

where r is the number of deleted firms. In this graph, we look for isolated low points; the r with isolated low points gives an indication of r outliers.

5.13.4 Finding outliers in R

The above calculations can easily be done in R. The following lines of R code provide a straightforward way to calculate the determinants D and $D^{(i)}$ and the ratio $R^{(i)}$. To illustrate the calculations, we use the data set `charnes1981.csv` with 5 inputs and 3 outputs available in the Benchmarking package.

```
c81 <- read.csv("charnes1981.csv")
x <- with(c81, cbind(x1,x2,x3,x4,x5))
y <- with(c81, cbind(y1,y2,y3))
xy <- cbind(x,y)
D <- det(t(xy)%*%xy)
i <- c(3,17) # firms to remove
xyi = xy[-i,]
Di <- det( t(xyi) %*% xyi )
Ri <- Di/D
```

The above calculation corresponds to the removal of just one group of two firms, namely firms 3 and 17. There are 70 firms in the data set, and therefore, there are $\binom{70}{2} = \frac{70!}{(70-2)!\,2!} = 2415$ different ways to remove 2 out of 70 firms. Thus, the above calculations should be done 2415 times simply to examine the possible impact of eliminating two firms from the total of 70 firms. Examining other groups sizes would soon make this direct approach impractical because the calculations would become too massive. Instead, we recommend the use of the function `ap` from the package FEAR. This function is coded in Fortran and unfortunately is only available for Windows and Linux, but it is stunningly fast. The following lines show how to use `ap`.

```
1 library(FEAR)
2 library(Benchmarking)
3 c81 <- read.csv("charnes1981.csv")
4 x <- with(c81, rbind(x1,x2,x3,x4,x5))
5 y <- with(c81, rbind(y1,y2,y3))
6 tap <- ap(X=x,Y=y,NDEL=12)
7 print(cbind(tap$imat,tap$r0),  na.print="", digit=2)
8 outlier.ap.plot(tap$ratio)
```

Line 1 loads the FEAR package in R, line 3 reads the data, lines 4 and 5 create input and output matrices in which the number of rows corresponds to the number of goods and the number of columns to the number of firms—note that input and output matrices for the FEAR functions are transposed matrices compared to standard use in R; we therefore use `rbind` instead of `cbind` . Line 6 calculates the minimum value of the Rs when deleting up to `NDEL` firms simultaneously and saves the result in the variable `tap`, line 7 prints the results as found in Table 5.7, and line 8 plots the log ratios in Fig. 5.4—this could also by done using the FEAR function `ap.plot(RATIO=tap$ratio)`, in which case the plot would look a little different. The details of the `ap` and `ap.plot` commands can be found in the FEAR manual pages.

The rows in Table 5.7 show which deletions give the minimum value of $R^{(r)}$; this minimum value is also shown. Thus, the first row, $r = 1$, shows that deleting

Table 5.7 The r removed observations corresponding to a minimum value of $R^{(r)}$ for the Charnes et al (1981) data

r	Deleted observations												$R^{(r)}_{\min}$
1	44												.4705
2	59	44											.2188
3	33	59	44										.1311
4	35	33	59	44									.0807
5	35	66	33	59	44								.0512
6	67	35	66	33	59	44							.0332
7	67	68	35	66	33	59	44						.0227
8	50	67	54	35	66	33	59	44					.0153
9	1	50	67	54	35	66	33	59	44				.0095
10	10	1	50	67	54	35	66	33	59	44			.0061
11	10	1	50	67	68	54	35	66	33	59	44		.0040
12	10	52	1	50	8	67	54	35	66	33	59	44	.0027

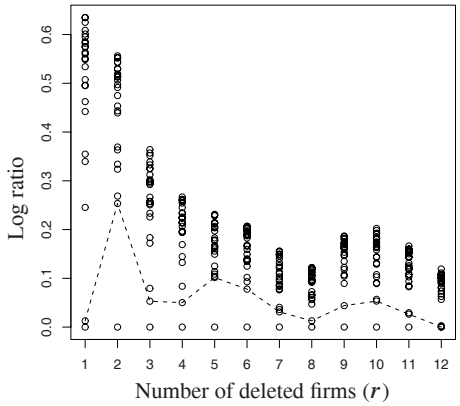


Fig. 5.4 Log–Ratio Plot for the Charnes et al (1981) data

firm 44 from the data set results in a value of $R^{(1)}$ at 0.4705 and that this value is the minimum value of $R^{(1)}$, i.e. the minimum R when just one firm is deleted from the dataset. The second row, $r = 2$, shows that deleting firms 59 and 44 gives a $R^{(2)}$ value of 0.2188; this is the minimum R value when two firms are deleted simultaneously. The same applies to the other rows.

To get a clear view of the minimum Rs and how they depend on the number of simultaneously deleted firms, we can look at Fig. 5.4, derived from line 8 in the above R listing of commands. In Fig. 5.4 we have plotted the ordered pairs $\left(r, \log\left(\frac{R^{(r)}}{R^{(r)}_{\min}}\right)\right)$. To look for outliers, we look for points in the graph where there is a gap between the points above 0 and the point at 0. A dashed line is drawn between

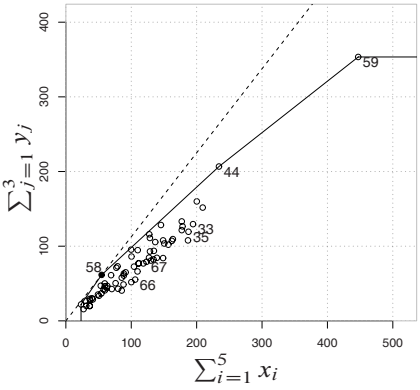


Fig. 5.5 Frontier for the Charnes et al (1981) data, aggregate input and output. The firm number is shown for the first 6 outlier firms and for the firm supporting the CRS line in the figure.

the points just above 0. We therefore have a group of outliers where the dashed line is far above 0.

In Fig. 5.4 we can see that the dashed line peaks at 2 deleted firms, $r = 2$; i.e. we have 2 outliers. From Table 5.7, we can see that these are firms 44 and 59. In Fig. 5.5 we have graphed the technology for aggregate input and aggregate output, which we have derived by simple summing inputs and of outputs. In this figure, one can easily see that the outliers are firms 44 and 59.

Firm 58, a firm on the frontier in all the three models (VRS, DRS, and CRS) marked with a filled circle in the frontier plot, is not considered an outlier based on the above process, not even when we consider groups of 12 outliers. Thus, a firm on the frontier is not necessarily an outlier if it is relatively close to other firms. The outlier firms on the frontier, the fully efficient firms, are thus firms that are different with regard to input and output size. The next peak in Fig. 5.4 shows that there might also be a group of 5-6 firms that could be considered outliers. Here, outlier firms are also among the worst firms, i.e. firms with low efficiency. They are numbered on the frontier plot.

Chapter 6
Statistical Analysis in DEA

6.1 Introduction

DEA is often classified as a non-statistical or deterministic approach that does not easily allow genuine hypothesis testing. Although DEA has not historically emphasized the use of traditional statistical tests, considerable progress has been made in this respect over the last 15 years. We will cover some important results in this chapter.

Initially, however, let us note that the background for DEA is operations research and management science. Management science is concerned with use of scientific, mostly mathematical, methods to solve real problems. This means that DEA studies have emphasized model-building as emphatically as they have model testing. That is, a DEA model developed for evaluation purposes is not to be evaluated solely based on its ability to explain and predict data in the best possible way. Basic properties of production economics like free disposability, economies of scale and convexity, the logic of the production structure from an engineering perspective, the relevance of the identified peers to industry representatives, etc., serves to validate the model just as statistical tests serve to validate a statistical model developed to replicate some underlying data generation process as closely as possible. Therefore, we maintain that interesting insights can arise from the use of DEA models without in the heavy use of statistical testing.

There are, of course, particular situations for which we are interested in performing hypothesis tests and constructing confidence intervals based on DEA models. Thus for example, we might wish to

- Test model-building assumptions like the returns to scale assumption
- Test for relevant and irrelevant inputs and outputs
- Test for differences between different groups of firms in terms of efficiency
- Test allocative and scale efficiency of a group of firms
- Test whether efficiency depends on external factors

In general, there are three ways to conduct such tests.

P. Bogetoft and L. Otto, *Benchmarking with DEA, SFA, and R*, International Series in Operations Research & Management Science 157, DOI 10.1007/978-1-4419-7961-2_6,

One is to rely on general *non-parametric tests*, i.e. tests used when the underlying distribution is unknown. We discuss some of these, including Kolmogorov–Smirnov tests and Kruskal–Wallis tests.

Another way is to rely on *parametric tests*, making assumptions regarding the underlying distribution of inefficiency and noise in the data. We will cover a series of such tests based on *asymptotic statistical theory*. Relying on asymptotic theory means that the theoretical properties are only established for large samples. However, simulation studies based on samples of moderate size, those including 50 firms and above, do suggest that they can be used quite generally.

The third approach, and one that has become popular with the development of effective computer programs, is the use of the *bootstrap*. The bootstrap is a computer-based method that can answer many statistical questions. The approach replicates sampling uncertainty by creating repeated samples of the original sample. We will spend most of this chapter covering bootstrap-based inference in DEA models.

In the appendix, we discuss the use of statistical methods in second-stage analyses, i.e. analyses performed after the development of a benchmarking model, to validate the model and to explore the possible causes of the variations in efficiencies. A common approach in such studies is tobit regression, and such analyses are not only relevant for DEA based benchmarking.

6.2 Asymptotic tests

In this section, we will assume that firm's efficiency is the realization of a random variable and that this is the sole reason why observed performance deviates from the underlying production possibility frontier; i.e. all deviations are efficiency-related, and there is no noise in the data.

Specifically, let us consider a DEA setting and assume that the true Farrell output efficiency ϕ, i.e.

$$\phi = \max\{ F \mid (x, Fy) \in T \}$$

is a random variable with values in $[1, \infty[$ and a density function g. Also, we assume that there is a non-zero likelihood of nearly efficient performance; i.e. $\int_1^{1+\delta} g(\phi)\, d\phi > 0$ for all $\delta > 0$.

In the following, it is importing to note that we distinguish between the true but unknown and unobservable technology T and a DEA estimate T^* of T. Now, it is clear that the estimated efficiency F in any finite sample of firms

$$F = \max\{ F \mid (x, Fy) \in T_\gamma^* \}$$

where

$$T_\gamma^* = \Big\{ (x,y) \in \mathbb{R}_+^m \times \mathbb{R}_+^n \mid x \geq \sum_{k=1}^{K} \lambda^k x^k,\ y \leq \sum_{k=1}^{K} \lambda^k y^k,\ (\lambda^1,\ldots,\lambda^K) \in \Lambda^K(\gamma) \Big\}$$

is biased downwards; i.e. it is always weakly smaller than true (in)efficiency ϕ, $F \leq \phi$. Recall here that $\Lambda^K(\gamma)$ is the restrictions on λ that depends on the returns to scale assumptions, i.e. fdh, vrs, or crs, as discussed in Sect. 4.4. The reason is that we have only observed a subset of practices, not necessarily the best practices, and the estimate of T^* of T is therefore an inner approximation, $T^* \subseteq T$, meaning that F measured against T^* is less than ϕ measured against T. Thus, estimated efficiency values never make a firm look less efficient than it really is, only more so. DEA-based estimates in this setting are cautious and puts the firms in a positive light.

However, asymptotically (with the number of firms going to infinity), this bias reduces to zero; that is, *the DEA estimators are consistent*. This holds as soon as the probability of observing nearly efficient firms is strictly positive, as we assumed above. Consistency is a nice statistical property because it means that for large samples, our evaluation is correct.

Additionally, one can show that if the density function g is monotonously declining (i.e. $f' > f \Rightarrow g(f') \leq g(f)$), then the DEA estimator F is the *maximum likelihood estimator* for ϕ.

The consistency results indicate that for large samples of firms, the distribution of F is similar to the distribution of ϕ. Therefore, in a large sample, the distribution of a test statistic $t(F)$ will be similar to the distribution of $t(\phi)$, and the distribution of $t(\phi)$ can be found from the density g of ϕ. This technique can be used to construct a series of tests as we do in the subsections that follow.

6.2.1 Test for group differences

If the set of K firms is divided into two groups with K_1 and K_2 firms, $K = K_1 + K_2$, we may be interested in testing whether there are significant differences between the efficiencies of the two groups—note that we use K, K_1 and K_2 as both the number of firms and the set of firms. This procedure may be relevant if we aim to test whether one special ownership structure is more efficient than another, whether one particular treatment is more effective than another, whether a specific region offers more favorable conditions for firms than another, etc.

Letting the density of the distributions of the efficiencies in the different groups be g_1 and g_2, respectively, we seek to test

$$H_0 : g_1 = g_2 \text{ against } H_A : g_1 \neq g_2.$$

As mentioned before, the distributions of $t(F)$ and $t(\phi)$ are asymptotically the same. If $t(\phi)$ is exponentially distributed, a chi-square distribution with 2 degrees of freedom, then $\sum_{k=1}^{K} t(F^k)$ is asymptotically χ^2–distributed with $2K$ degrees of freedom.

Under the null hypothesis, the two groups have the same distribution of efficiency, and the ratio

$$T_{EX} = \frac{\sum_{k \in K_1} t(F^k)/K_1}{\sum_{k \in K_2} t(F^k)/K_2}$$

is the ratio of two asymptotically χ^2–distributions and is therefore asymptotically distributed as a Fisher distribution with $2K_1$ and $2K_2$ degrees of freedom, $T_{EX} \overset{a}{\sim} F(2K_1, 2K_2)$. Note that T_{EX} might be greater or less than 1 such that the test is two-sided.

If we assume that true efficiency is $\phi = 1 + \epsilon$ where ϵ is exponential distributed, then we should simply use $t(F) = F - 1$ such that

$$T_{EX} = \frac{\sum_{k \in K_1} (F^k - 1)/K_1}{\sum_{k \in K_2} (F^k - 1)/K_2}$$

and reject the hypothesis if T_{EX} is greater than the 95% quantile in the distribution $F(2K_1, 2K_2)$.

Likewise, if $t(\phi)$ has a half-normal distribution, then $t(\phi)^2$ is χ^2 distributed, and therefore, $\sum_{k=1}^{K} t(F_k)_2$ is asymptotically χ^2–distributed with K degrees of freedom. The test statistic

$$T_{HN} = \frac{\sum_{k \in K_1} t(F^k)^2/K_1}{\sum_{k \in K_2} t(F^k)^2/K_2}$$

is therefore distributed as $F(K_1, K_2)$. This will be the case if, for example, $\phi - 1$ has a half-normal distribution, and in this case, we should again use $t(F) = F - 1$.

Lastly, if we have no a priori assumptions about the distribution of ϕ_1 and ϕ_2, we may use the non-parametric Kolmogorov– Smirnov test statistic

$$T_{KS} = \max_{k=1,\ldots,K} \left\{ |G_1(F^k) - G_2(F^k)| \right\}$$

where G_1 and G_2 are the empirical cumulative distributions in the two subsets such that T_{KS} is the largest vertical distance between the cumulative distributions. Large values of T_{KS} as evaluated via the Kolmogorov-Smirnov test as an indication that H_0 is false. Note that this test depends on the rank (i.e. the order) of F^k only and not on the individual values of F^k.

Another non-parametric test based on ranks is the Kruskal–Wallis test used to test groups of data. We will not show how to run this test but would like to note that the test only depends on the rank of the observations. This test is helpful because it can be used to test the hypothesis that several groups have the same distribution.

Numerical example in R: Milk producers

We want to test data from a group of milk producers to determine if efficiency depends on the breed of cow. The inputs are cost categories, and the output is milk. Group 1 is comprised of farmers without jersey cows, whereas group 2 is comprised of farmers with jersey cows.

Implementing the T_{EX} and T_{HN} tests in R is easy; these tests are simply a matter of summing the efficiencies with 1 subtracted. The commands `qf` and `pf` calculate the quantile (.95 for 95% or 5% tail probability) and the probability in the Fisher distribution. The calculated output efficiencies are split into two groups `F1` and `F2` based on the value of the two-level factor `race`, and the test evaluates whether the efficiency of the two groups is identical.

The Kolmogorov– Smirnov and the Kruskal–Wallis tests are more complicated, but R already contains special methods for those tests; therefore, it is easy to use them in R.

The code and output for the tests are shown here:

```
> library(Benchmarking)
> cattle = read.csv("projekt.csv")
> attach(cattle)
> kgMilk <- milkPerCow * cows
> x <- cbind(unitCost, capCost, fixedCost)
> y <- matrix(kgMilk)
> FF <- eff(dea(x,y,ORIENTATION="out"))
> TEX <- sum(F1-1)/length(F1) / (sum(F2-1)/length(F2))
> TEX
[1] 1.989044
> qf(.025, 2*length(F1), 2*length(F2))
[1] 0.6369572
> qf(.975, 2*length(F1), 2*length(F2))
[1] 1.682756
> pf(TEX, 2*length(F1), 2*length(F2))
[1] 0.9947547
> THN <- sum((F1-1)^2)/length(F1) / (sum((F2-1)^2)/length(F2))
> THN
[1] 2.000593
> qf(.025, length(F1), length(F2))
[1] 0.5357977
> qf(.975, length(F1), length(F2))
[1] 2.148472
> pf(THN, length(F1), length(F2))
[1] 0.9628421
> # Kolmogorov-Smirnov test
> ks.test(F1, F2)

        Two-sample Kolmogorov-Smirnov test
data:  F1 and F2
D = 0.4893, p-value = 0.0006954
alternative hypothesis: two-sided

> # Kruskal--Wallis, 2 groups
```

```
> kruskal.test(FF, race=="jersey")

        Kruskal-Wallis rank sum test
data:  FF and race == "jersey"
Kruskal-Wallis chi-squared = 11.6309, df = 1, p-value = 0.0006487
```

The value of the T_{EX} is calculated to be 1.989044, and as the 97.5% upper critical value (the size of the test is 5%) in the F distribution with 80 (the number of firms in group 1) and 21 (the number of firms in group 2; breed "jersey") is 1.68, we reject the hypothesis that the distribution of efficiency in the groups is identical. The T_{HN}, on the other hand, is 2.00, and the upper critical value is 2.148. Thus, we do not reject the hypothesis that they are identical; rather, the groups could be identical. The results of both the Kolmogorow–Smirnov test and the Kruskal–Wallis test lead us to reject the null hypothesis. Based on the boxplot and densities in Fig. 6.1, it

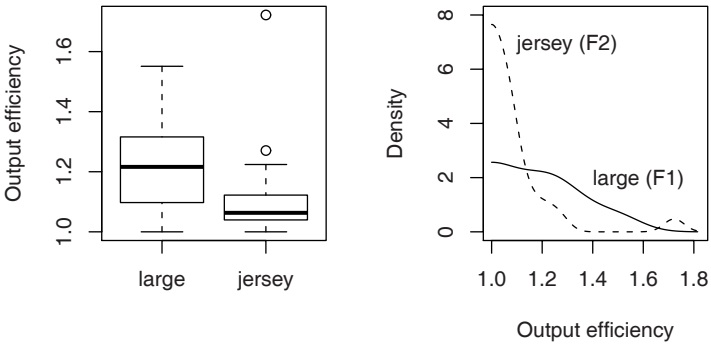


Fig. 6.1 Boxplot and densities for output efficiency of the two subgroups

does look as if group 2 (the "jersey" breed) has steeper density and mass closer to 1 than group 1. Most of our tests also show that the difference is significant, and what we see in the figure is therefore most likely not a matter of chance. One result that emerges is that the for group F2 ("jersey"), the average output efficiency is lower than that for group F1 ("large"); i.e. F2 is more efficient than F1 on average. Note that there is an outlier in group F2, indicated both at the top of the boxplot as a circle and in the density illustration as a blip to the far right.

6.2.2 Test of model assumptions

In model development and model validation, we may want to test if an alternative model specification better represents firm performances. We might, for example, be interested in testing whether we can assume variable return to scale or whether some outputs can be eliminated from the model specification.

Here we will distinguish not between two groups of observations, but rather distinguish between two sets of model assumptions, or what accounts to the same, distinguish between two technology sets. In Sect. 4.3, we argued that the estimated technology set should be the smallest set containing the data and fulfilling certain assumptions (the minimal extrapolation principle). The question we ask here is therefore whether an estimated technology set can be made even smaller by adding further restrictions and still be in agreement with data. Let the technology set be T_1, and let the smaller technology set be T_2. For example, we could have the same assumptions regarding the technology sets but assume CRS in T_1 and VRS in T_2, with the additional restriction $\sum_{k=1}^{K} \lambda^k = 1$. Likewise, the technology set T_1 could include n outputs, and the technology set T_2 could include $n' > n$ outputs; the greater number of outputs would result in the existence of more restrictions and therefore yield a smaller technology set.

In both examples, and in general, a smaller technology set (i.e. one with more restrictions) results in a better (or unchanged) efficiency level; for input efficiency, we obtain $E_1 \leq E_2 \leq 1$, and for output efficiency, we obtain $F_1 \geq F_2 \geq 1$ where the subscript of the efficiencies E and F is a product of the corresponding technology set T_1 and the smaller set T_2.

In statistical language, technology set T_1 represents the null hypothesis and the smaller technology set T_2 the alternative. We test technology hypothesis T_1 against alternative T_2.

If the efficiencies calculated under T_1 are very different from the efficiencies calculated under T_2, the two technologies are not at all similar, and we should reject the null-hypothesis technology T_1 and opt for the alternative technology T_2; the extra restrictions in T_2 are of real importance. If the efficiencies are more or less the same, then the extra restrictions are of no importance, and we opt for the null-hypothesis technology T_1. Therefore, we can test the technology assumptions by testing whether efficiency is the same under the two technologies.

Now, let the distribution of the efficiency scores for K firms under the two technology assumptions T_1 and T_2 be g_1 and g_2, respectively. We will then test the hypothesis

$$H_0 : g_1 = g_2 \text{ against } H_A : g_1 \neq g_2$$

using the same ideas as above, except that we now sum the figures for all firms in both the numerator and the denominator. If we accept the hypothesis H_0, we use technology T_1, whereas if we reject the hypothesis, we use technology M_2. More specifically, if $t(\phi_1)$ and $t(\phi_2)$ are exponentially distributed for some monotone transformation $t(\cdot)$, then just as before, the test statistic

$$T_{EX} = \frac{\sum_{k=1}^{K} t(F_1^k)}{\sum_{k=1}^{K} t(F_2^k)},$$

where F_1^k and F_2^k are the output efficiency of firm k based on technologies T_1 and T_2, respectively, will follow a F–distribution under H_0 with $2K$ and $2K$ degrees of freedom, $F(2K, 2K)$.

The test is one-sided as $T_{EX} \geq 1$, and therefore, the critical value for a test of size 5% is the 95% quantile in the F–distribution with $2K$ and $2K$ degrees of freedom, $F(2K, 2K)$; i.e. for large values of T_{EX}, we reject the null hypothesis H_0 that model M^1 is true.

Likewise, if $t(\phi_1)$ and $t(\phi_2)$ have a half-normal distribution for some monotone transformation $t(\cdot)$, then we can use the test statistic

$$T_{HN} = \frac{\sum_{k=1}^{K} t(F_1^k)^2}{\sum_{k=1}^{K} t(F_2^k)^2}$$

with large values in a $F(K, K)$ distribution as critical values for the test of H_0.

Lastly, if we have no a priori assumptions about the distribution of ϕ_1 and ϕ_2, we can use the non-parametric Kolmogorov– Smirnov test statistic

$$T_{KS} = \max_{k=1,\ldots,K} \{ |G_1(F^k) - G_2(F^k)| \}$$

where G_1 and G_2 are the empirical cumulative distributions in the two models such that T_{KS} is the largest vertical distance between the cumulative distributions. Large values for T_{KS} indicate that the distributions differ and therefore that H_0 is false; the null hypothesis H_0 is rejected.

Numerical example in R: Milk producers

Implementing the tests for model assumptions is just as easy as implementing the tests of group differences. However, we present an example anyway to introduce yet another example of a hypothesis.

So far, we have used two examples to test our model assumptions. Here, we use a third example to test whether to include fewer inputs. The null hypothesis is technology T_1 with m inputs, whereas the alternative is technology T_2 with $m' > m$ inputs. Again, the alternative includes more restrictions and specifically more input restrictions in the LP formulation. The test statistics are as previously described.

We use the same data set that we used to test group differences. We want to test whether we really need capacity costs when we already include the number of cows and whether veterinary expenses are important on their own even though they are part of unit costs. Thus, the alternative technology set T_2 includes among its inputs the number of cows and veterinary expenses, whereas technology T_2, the null hypothesis, excludes these two inputs.

The input matrix `x1` in the example below excludes the variables in question, whereas the input matrix `x2` includes them. The following code reads data, calculates efficiency and makes graphs as shown in Fig. 6.2. The graphs are slightly different from the ones we presented in the test for group differences.

```
library(Benchmarking)
cattle = read.csv("projekt.csv")
kgMilk <- with(cattle, milkPerCow * cows )
```

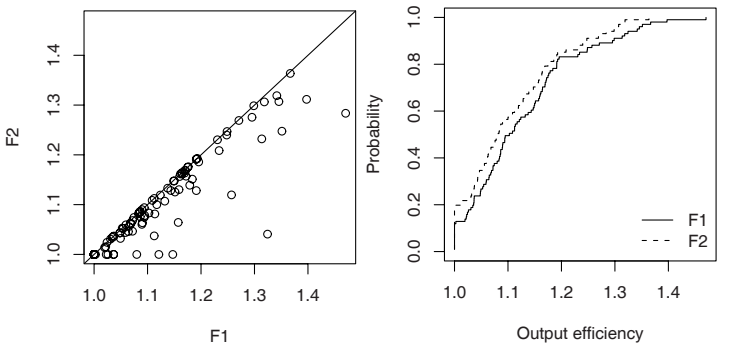


Fig. 6.2 Efficiency when capacity cost and veterinary costs are excluded (F_1) and included (F_2) in the inputs for milk production: comparing efficiencies and the empirical distribution of efficiencies.

```
x1 <- with(cattle, cbind(unitCost,          fixedCost,     cows))
x2 <- with(cattle, cbind(unitCost, capCost, fixedCost, vet, cows))
y <- matrix(kgMilk)
F1 <- eff(dea(x1,y,ORIENTATION="out"))
F2 <- eff(dea(x2,y,ORIENTATION="out"))

plot(F1,F2, xlim=range(F1,F2), ylim=range(F1,F2))
abline(0,1)

K <- length(F1)
plot(sort(F1), (1:K)/K, type="s", ylim=c(0,1),
      ylab="Probability", xlab="Output␣efficiency")
lines(sort(F2), (1:K)/K, type="s", lty="dashed")
legend("bottomright",c("F1","F2"),
      lty=c("solid","dashed"),bty="n")
```

The box plot shows that the two technologies T_1 and T_2 are only slightly different in terms of efficiency; the spread is slightly greater for `F1` that for `F2`. The same pattern is seen in the top right plot, where some of the efficiencies are identical (i.e. on the diagonal line) and some for `F1` are larger than those for `F2` (below the diagonal line). This is no surprise given that the number of inputs is smaller in `F1`; firms will have unchanged or greater output efficiency, as discussed in Sect. 4.6 on page 93. The bottom figure shows the empirical distribution. The distribution of `F2` is above that of `F1`; for every level of efficiency, the proportion of firms at that level or lower is larger for technology T_2 than for technology T_1.

The problem is whether the difference that we see is statistically significant. This is where the test statistics come into play. Based on the above calculations for the two efficiencies, the test statistics are calculated below.

```
> TEX <- sum(F1-1)/length(F1) / (sum(F2-1)/length(F2))
> TEX
[1] 1.211835
> qf(.95, 2*length(F1), 2*length(F2))
[1] 1.261131
> pf(TEX, 2*length(F1), 2*length(F2))
```

```
[1] 0.9135035
>
> THN <- sum((F1-1)^2)/length(F1) / (sum((F2-1)^2)/length(F2))
> THN
[1] 1.381849
> qf(.95, length(F1), length(F2))
[1] 1.389417
> pf(THN, length(F1), length(F2))
[1] 0.9471316
>
> # Kolmogorov-Smirnov test
> ks.test(F1, F2, alternative = "greater")

        Two-sample Kolmogorov-Smirnov test

data:  F1 and F2
D^+ = 0, p-value = 1
alternative hypothesis: the CDF of x lies above that of y

Warning message:
In ks.test(F1, F2, alternative = "greater") :
  cannot compute correct p-values with ties
> # Kruskal--Wallis
> kruskal.test(list(F1, F2))

        Kruskal-Wallis rank sum test

data:  list(F1, F2)
Kruskal-Wallis chi-squared = 2.519, df = 1, p-value = 0.1125
```

The T_{EX} and T_{HN} are estimated to be 1.21 and 1.38, and both fall below the critical value, the 95%–quantile. The results of the Kolmogorof–Smirnof test and the Kruskal–Wallis test both support the same conclusion. Note that the probabilities for these tests are tail probabilities. Therefore, we do not reject the null hypothesis that we need to include capacity cost and veterinary costs among the inputs, and for all uses of the technology, we should be using T^1 with the fewest input variables.

Practical application: DSO regulation

In the regulation of German electricity distribution operators, DSOs, a series of tests were undertaken to ensure that models did not unintentionally favor or disadvantage specific types of companies. We will discuss regulation in greater detail in Chap. 10. The tests for the DSO technologies was conducted as second-stage tests of the best of four scores that the regulation prescribed using non-parametric Kruskal–Wallis tests, cf. also Chap. 10. However, we could also have used tests like those above to directly evaluate the individual DEA models and test for the impact of such factors as 1) whether the DSO is located in what was formerly West or East Germany or 2) whether the DSO is also involved in gas distribution, water distribution etc.

The same regulations also stipulate that no single DSO can have too large an impact on average efficiency in the DEA models. This requirement was tested using the test statistic

$$\frac{\sum_{h\in K\setminus k}\left(E(h,K\setminus k)-1\right)^2}{\sum_{h\in K\setminus k}\left(E(h,K)-1\right)^2}.$$

Here, K is both the set and the number of DSOs in the data set, and k is a potential outlier. Also, $E(h,K)$ is the efficiency of h when all DSOs are used to estimate the technology, and $E(h,K\setminus k)$ is the efficiency when DSO k does not enter into the estimation. The test therefore compares the average efficiency of the other operators when DSO k cannot affect the technology with the average efficiency of the other DSOs when DSO k is part of the evaluation process. Because $E(h,K\setminus k)\geq E(h,K)$, this ratio is always less than or equal to 1, and the smaller the ratio, the larger the impact of k; i.e. small values will be an indication that k is an outlier. We see that this line of thought resembles the model specification test problems above, which suggests that we can evaluate the test statistic in a $F(K-1,K-1)$ distribution.

6.3 The bootstrap method

Bootstrap is a general computer-based statistical method for calculating the accuracy of statistical estimates. Generally, "pulling oneself up by one's bootstraps" means to succeed based on one's own efforts despite very difficult circumstances and without help from anyone. The statistical bootstrap method has some of this flavor and recalls the story of Baron von Munchausen, who pulled himself and his horse out of a swamp by pulling on his own hair while holding on to the horse with his legs. In the following pages, we first give a short introduction to bootstrap as a general method and then explore the details of bootstrap DEA models.

The basic idea of bootstrap is to sample observations with replacements from one's data set and thereby create a new "random" data set of the same size as the original. Using this dataset, one can calculate the necessary statistics, called replicates. This process is repeated to create a *sample of replicates*. Based on this sample, we can draw conclusions about the distribution of the statistics in which we are interested.

Let us consider a very simple example, a sample of n observations $x_1, x_2, \ldots, x_n$. Imagine that we have observed 7 numbers 94, 197, 16, 38, 99, 141, and 23. The mean is $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 86.86$, and the (unbiased) standard error is $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2} = 66.77$. The estimate of the standard error of the mean is $\frac{s}{\sqrt{n}} = 25.24$. The standard error is very easy to estimate when we simply wish to determine the variance of the mean because we can use an explicit formula. Unfortunately, we do not always have an explicit formula for the standard error or for variance.

Table 6.1 The bootstrap algorithm for estimating standard errors

1. Select B independent bootstrap samples $x^1, x^2, \ldots, x^B$, i.e. a sample drawn with replacement from our data set.
2. Calculate the estimate for each bootstrap sample:

$$t(x^b) \quad (b = 1, \ldots, B) .$$

3. Estimate the standard error using the sample standard error of the B replications

$$\hat{s}_B = \sqrt{\frac{1}{B-1} \sum_{b=1}^{B} \left(t(x^b) - \bar{t}\right)^2}$$

where $\bar{t} = \frac{1}{B} \sum_{b=1}^{B} t(x^b)$.

If instead of investigating the mean we wish to find the median and the variance of the median, we must undertake a much more complicated process because the formula for calculating the variance of the median is not easily determined. This is where the bootstrap method becomes key.

A bootstrap sample in this case is a random sample obtained by sampling 7 (the number of elements in the sample) elements or data points *with* replacements from our original sample. Hence, the bootstrap sample could be $x^b = (x_6, x_1, x_4, x_1, x_3, x_3, x_5)$, i.e. 141, 94, 38, 94, 16, 16, and 99. Based on this bootstrap sample, we estimate the statistic $t(x^b)$ we are interested in: here, the median. Now, instead of trying to calculate the standard deviation of the estimated median, we make B bootstrap replications. For each bootstrap replication b, we calculate $t(x^b)$, the median. As the bootstrap estimate of the standard error of $t(x)$ with B replications, we use $\hat{s}_B = \sqrt{\frac{1}{B-1} \sum_{b=1}^{B} (t(x^b) - \bar{t}\,)^2}$ where $\bar{t} = \frac{1}{B} \sum_{b=1}^{B} t(x^b)$ is the mean over the replications of the statistic we are interested in.

The idea of the bootstrap method is that if the empirical distribution of x^b corresponds more or less to the true distribution of x, then the empirical distribution of $t(x^b)$ will correspond more or less to the true distribution of $t(x)$. This means that we can use the empirical distribution of $t(x^b)$ as the true but unknown distribution of $t(x)$. Thus, when we are interested in the variance of the median, $t(x)$, which is difficult or impossible to determine, we can simply use the empirical variance of the median of the bootstrap, $t(x^b)$, which is much easier to obtain.

The bootstrap method can be described as the algorithm in Table 6.1. The limit of $\hat{s}_B$ as B goes to infinity is the ideal bootstrap estimate.

Luckily, we do not have to program the algorithm in Table 6.1 ourselves; it is part of the package boot in R, and now we show how to use it in the small numerical example we have just seen.

Table 6.2 Bootstraping the variance of the median in a sample with 7 numbers

```
library(boot)
treat <- c(94, 197, 16, 38, 99, 141, 23)
func <- function(d,i) { median(d[i]) }
B <- 200
boo <- boot(treat, func, B)
sqrt(var(boo$t))
mean(boo$t)
hist(boo$t,main=NULL)
```

Numerical example in R

Bootstrap is easy in R because the package boot contains the function `boot`, which organizes the resampling and calculation of a statistic (function) we provide; this is just an implementation of the algorithm in Table 6.1. In our example in which we investigate the variance of the median, we use the R script in Table 6.2. The first line is the command to load the library `boot` that contains the commands and methods for bootstrap in R. The second line defines our data set, our original sample, as the variable `treat`. To use the R function `boot`, we must define a function that calculates the statistic of interest. In our case the function must calculate the median, and it must be defined with two arguments, the first the original data and the second a vector of indices, frequencies or weights that define the bootstrap sample. Here, the function is called `func`, and the two arguments are `d` for data and `i` for the indices, such that `d[i]` is a bootstrap sample and the return of the function is the median of the bootstrap sample `d[i]`. Next, we define variable `B` as the number of bootstrap replicates; in this case, we use 200 replicates. To actually generate the bootstrap replicates, we use the R function `boot`. This function takes 3 arguments: the original sample, the function we have defined to calculate the statistics of interest, and the number of replicates (bootstrap iterations) we seek, here the defined by the variable `B`.

The function `boot` can take many more arguments than we use here; see the manual, `>?boot`, for others.

The output from the bootstrap function is put into the variable `boo`, a boot object. Hereafter, we can gain access to the replicates of the 200 calculated statistics (medians in our case) in the component `t` in the object/variable `boo`, i.e. the variable `boo$t`. Now we can easily calculate the variance of the median as `boo$t`, and if we want to determine the standard error, we can simply take the square root. The resulting standard error of the median of our sample `treat` is

```
> sqrt(var(boo$t))
  38.00217
```

showing that the standard error of the median of our 7 numbers is 38. A histogram of the bootstrap replicas is shown in Fig. 6.3. The figure indicates that the most common median is between 90 and 99, and based on the data set, we can see that

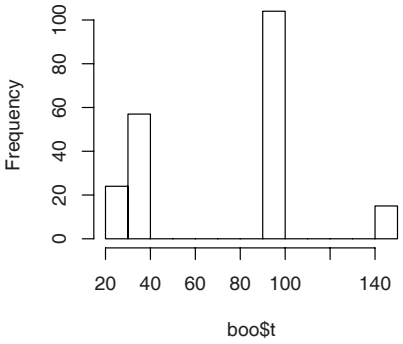


Fig. 6.3 Histogram of bootstrap replicas for the median of the 7 numbers

Table 6.3 Bootstrap the median of numbers with different replications in R

```
library(boot)
func <- function(d,i) { median(d[i]) }
treat <- c(94, 197, 16, 38, 99, 141, 23)
Ber <- c(10,50,100,250,500,1000,5000,10000,1000000)
res <- NULL
for(B in Ber) {
   boo <- boot(treat, func, B)
   res <- c(res, format(sqrt(var(boo$t)), digits=3))
}
Ber     # print Ber
res     # print res, the results
rbind(Ber, res)
```

it must be 94 or 99—the median in the original data set `treat` is 94. The second most common median is just below 50 and the actual number is 38.

If we make the same calculations again, we may obtain a figure for variance that is somewhat different because we obtain another series of replications. However, if the number of replications is very large, then each time we repeat the bootstrap series of replications, the variance will be almost the same. The question is then how many replications we should conduct to develop a stable estimate of the variance?

The calculated standard errors of the median from several bootstraps when the number of bootstrap replicates B is ranging from 10 to 1 000 000 is calculated using the R program in Table 6.3. The results achieved by running this code are shown in Table 6.4; we have run the program several times and show the different standard errors in the different rows. When the number of bootstrap replications is larger than 1000, there is hardly any difference between the levels of variance for the different runs. Thus, the desired level of precision of the estimated variance determines the number of replications.

For a bootstrap sample of size 10, one of the standard errors differs substantially from the other bootstrap samples, as can be seen in Table 6.4. Based on considerations like this one, it is suggested in the literature that bootstrap samples, B, ranging

Table 6.4 Bootstrap estimates of standard error of the median

B:	10	50	100	250	500	1 000	5 000	10 000	1 000 000
Run 1: Std.err:	32.7	38.5	37.5	38.9	36.7	38.3	37.9	38.1	37.8
Run 2: Std.err:	38.7	44.5	40.4	39.7	38.4	37.9	37.8	37.4	37.8
Run 3: Std.err:	2.58	43.3	33.2	37.0	36.8	38.2	37.7	37.8	37.8
Run 4: Std.err:	35.9	37.8	41	37.3	38.6	38.7	37.6	38.0	37.9

from 50 to 200 usually make the bootstrap a good standard error estimator. As we shall see later, however, these suggested numbers of bootstrap replications are to small for DEA models.

If we want to find the variance of another function or statistic instead of the median of our sample, we can simply redefine the function `func` to calculate the new statistic, which may include very complicated calculations (as is the cace, for instance, with DEA efficiency). If we want to consider another sample, we can just change the contents of `treat`.

6.3.1 Confidence interval

Using the bootstrap sample, we can also directly determine the confidence intervals for the statistic. This approach yields more precise results than do efforts to construct the confidence intervals based on the estimated standard deviation because the latter technique rests on the assumption that the distribution in question is symmetric and can be reasonably approximated using a normal distribution. This is not the case for the aforementioned example intended to determine the median of the 7 numbers.

To find a 50% confidence interval for the sample, we can use the command quantile in R, as shown in Table 6.5. The results are shown in Fig. 6.4. In the figure, the

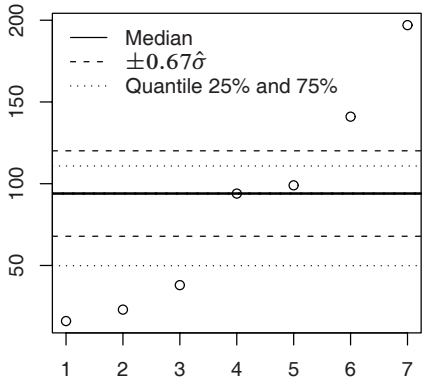


Fig. 6.4 Confidence interval for median of 7 numbers based on 200 replicats

Table 6.5 Calculating a 50% confidence interval for the median of 7 numbers

```
library(boot)
treat <- c(94, 197, 16, 38, 99, 141, 23)
func <- function(d,i) { median(d[i]) }
B <- 200
boo <- boot(treat, func, B)
sqrt(var(boo$t))
mean(boo$t)

quant <- .50 # 50% confidence interval
ci <- boot.ci(boo, conf=quant)
m <- mean(boo$t)
b <- m - median(treat)    # bias
mu <- m -b                # bias corrected median
sd <- sqrt(var(boo$t))    # std.error
quantile(boo$t,c((1-quant)/2, 1-(1-quant)/2) ) -b
```

7 numbers are shown in sorted order, and the median is marked with the solid line through the point at 94. The 50% confidence interval based on a normal approximation is shown as a dashed line, and of course, it is symmetric around the median. The dotted line is based on the command `quantile`, and this confidence interval is not symmetric around the median. The upper line is a little lower that the normal line, and the lower line is much lower than the normal line. This corresponds to the histogram in Fig. 6.3, where the distribution does not seem to be symmetric. Based on the actual numbers in the sample `treat`, the 50% interval for the median 94 is from 38 to 99. This corresponds to the histogram in which one can see that the median in half of the replicas is between 35 and 100.

6.4 Bootstrapping in DEA

We will now discuss how to estimate the variance of efficiency measures for a sample of firms using the bootstrap method. Let the observations be $(x^1, y^1), \ldots, (x^K, y^K)$ and the corresponding Farrell input efficiency measures be $E^1, \ldots, E^K$, i.e. $E^k = \min\{\, \theta \in \mathbb{R}_+ \mid (\theta x^k, y^k) \in T \,\}$. None of what follows would change if we considered Farrell ouput efficiency instead.

It does not make sense to compute variance as $\frac{1}{n-1}\sum_{k=1}^{K}(E_k - \bar{E})^2$ because then we would be assuming that all the firms have efficiencies based on a distribution with the same mean and therefore that all differences in efficiency are purely random and not systematic; firms with high efficiency would then be highly efficient by chance and because they are good at what they do.

Instead, we use our observations as a sample $\mathfrak{X} = \{(x^1, y^1), \ldots, (x^K, y^K)\}$ of inputs and outputs from K firms that we can use to estimate the technology set T via DEA assuming variable returns to scale (vrs)

$$\widehat{T} = \{\,(x, y) \mid x \geq \sum_{k=1}^{K} \lambda^k x^k, y \leq \sum_{k=1}^{K} \lambda^k y^k, \lambda^k \geq 0, \sum_{k=1}^{K} \lambda^k = 1\,\}.$$

The DEA estimated efficiency scores are then

$$\widehat{E}^k = \min\{\,\theta \in \mathbb{R} \mid (\theta x^k, y^k) \in \widehat{T}\,\} \qquad (k = 1, \ldots, n)$$

where we have used the estimated technology set $\widehat{T}$ for the technology set T.

We use this procedure to consider the sample $\mathcal{X} = \{(x^1, y^1), \ldots, (x^K, y^K)\}$ as a realization of identically and independently distributed random variables (X, Y) with a probability distribution $\mathcal{P}$ with support in T; i.e. we assume that there is no observational uncertainty in the sense that $(x^k, y^k) \in T$ with probability 1. In Chap. 7, we introduce a parametric method that allows for this form of observational uncertainty.

The distribution of $\widehat{E}^k$ and $\widehat{T}$ depends on the distribution of the sample of observations $\mathcal{X}$. However, this relationship is complex; the sample $\mathcal{X}$ is generated by the probability distribution $\mathcal{P}$, of which we have no direct knowledge. To derive a reasonable estimate $\mathcal{P}^*$ of $\mathcal{P}$, we can use the bootstrap, i.e. a sample with replacements from the original set of observations. Using this bootstrap estimate $\mathcal{P}^*$ of $\mathcal{P}$, we can generate a sample $\mathcal{X}^*$ from the distribution $\mathcal{P}^*$, then calculate a DEA estimate T^* for the technology and estimate efficiency as $E^{k*} = \min\{\,\theta \in \mathbb{R} \mid (\theta x^k, y^k) \in T^*\,\}$. When we repeat this sample generation process many times, we obtain many estimates of E^{k*} and can then calculate the empirical variance of E^k $(k = 1, \ldots, n)$.

6.4.1 Naive bootstrap

There are two ways to perform an ordinary bootstrap for the DEA model. Unfortunately as we will see, neither of them is satisfactory, and we will therefore present a better alternative.

The two simple but unsatisfactory methods are as follows:

1. Bootstrap the set directly $\{E^1, \ldots, E^K\}$ as we did in Sect. 6.3 on the variable `treat`. In using this method, we assume that all the E's are independent and identically distributed with a probability distribution $\mathcal{P}_E$. This implies that any differences in efficiency are purely random because they all come from the same distribution $\mathcal{P}_E$. On that basis, firm inefficiency appears to be related neither to x^k nor to y^k. This outcome is not satisfactory.
2. We bootstrap the set $\mathcal{X} = \{(x^1, y^1), \ldots, (x^K, y^K)\}$, and for each bootstrap sample, b, we estimate the technology T^b and the efficiency E^{kb} for firm k. When we make B bootstrap samples, B replicas, we can calculate the mean and variance of the efficiency of firm k using $\bar{E}^{k*} = \frac{1}{B}\sum_{b=1}^{B} E^{kb}$ and $\frac{1}{B}\sum_{b=1}^{B}(E^{kb} - \bar{E}^{k*})^2$.

One problem is that for some firm k, (x^k, y^k) may not be in a bootstrap sample, a replica b, and (x^k, y^k) may not be in the technology set generated by the bootstrap sample, $(x^k, y^k) \notin T^{*b}$. This implies that we have a firm outside the technology set, but one of our assumptions was that all observations are inside the technology set with probability 1. If we calculate the efficiency anyway, we find in this case that $E^{kb} > 1$.

This could easily happen for firms where $E^k = 1$ as a bootstrapped technology set T^* will always be a subset of the technology set $\widehat{T}$ estimated on all observations, $T^* \subset \widehat{T}$, and therefore $E^{k*} \geq E^k$. Essentially, we could in many bootstrap samples find firms where $E^{kb} > 1$.

We could disregard the requirement that all observations be inside the technology set and just use $E^{kb} = 1$ if we obtained $E^{kb} > 1$. One problem with this technique is that the probability of E near 1 will be underestimated because the method puts a positive probability mass at $E = 1$ and the estimated distribution is therefore not a good estimate of the empirical distribution near $E = 1$.

6.4.2 Smoothing

The bootstrap sample will nearly always contain repeated values, and if n is small, then it will even contain values repeated several times. To avoid spikes in the distribution like those that we saw in Fig. 6.3, it is advisable to use a *smoothed bootstrap* method to smoothe the distribution. As before, we want to bootstrap the sample $(x^1, \ldots, x^K)$. Here, the sample is constructed in the following way: For $r = 1, \ldots, K$

1. choose k at random with a replacement from $\{1, \ldots, K\}$,
2. generate ϵ from a standard normal distribution,
3. set $z^r = x^k + h\epsilon$ and call h the window or band width.

Our bootstrap sample is then $(z^1, \ldots, z^K)$, not a real sample from the original sample $(x^1, \ldots, x^K)$, but a smoothed sample. In this way, we smoothen the fixed number of points to imitate a continuous distribution function of the inputs x. The distribution for these smoothed points is a normal distribution with variance h^2 and is therefore symmetric around the observation points. When we use the bootstrap sample to calculate the efficiencies E, there might be a problem for efficiencies near the boundary at 1 because they must be equal to or below 1. To handle problems related to E near 1, we can use a reflection method, augmenting the dataset by adding reflections of all the points in the bootstrap; i.e. whenever we have efficiency E, we augment the dataset with the reflection on 1, $2 - E$, such that E and $2 - E$ are symmetric around 1. Then, we simply use the value below or equal to 1.

6.4.3 Bias and bias correction

In what follows we shall use the following terms:

θ^k	The true efficiency based on the true but unknown technology T
$\hat{\theta}^k$	DEA-estimated efficiency and $\widehat{T}$ the estimated DEA technology
θ^{kb}	The bootstrap replica b estimate based on the replica technology T^b
θ^{k*}	The bootstrap estimate of θ^k
$\tilde{\theta}^k$	The bias-corrected estimate of θ^k

The DEA estimate is upward biased: if there are no measurement errors, then all of the observations in the sample are from the technology set $\widehat{T} \subset T$. Then in $\hat{E}^k \geq E^k$, because we are minimizing over a smaller set (i.e. the estimated efficiency is an upward-biased estimate of E^k,), the estimated efficiency may be larger than the real efficiency. The size of $\widehat{T}$ depends on the sample, and therefore, E^k is sensitive to sampling variations in the obtained frontier. If there are measurement errors, then there is no direct subset relation between $\widehat{T}$ and T.

To eliminate the bias, we first estimate the bias and obtain a bias-corrected estimate. We can estimate the bias as

$$\text{bias}^k = \text{EV}(\hat{\theta}^k) - \theta^k.$$

Unfortunately, we do not know the distribution of θ^k, so we cannot calculate $\text{EV}(\hat{\theta}^k)$. This is where the bootstrap enters in. When θ^{kb} is a bootstrap replica estimate of θ^k, the bootstrap estimate of the bias is

$$\text{bias}^{k*} = \frac{1}{B}\sum_{b=1}^{B} \theta^{kb} - \hat{\theta}^k = \bar{\theta}^{k*} - \hat{\theta}^k.$$

A bias-corrected estimator of θ^k is then

$$\tilde{\theta}^k = \hat{\theta}^k - \text{bias}^{k*} = \hat{\theta}^k - \bar{\theta}^* + \hat{\theta}^k = 2\hat{\theta}^k - \bar{\theta}^{k*}.$$

The precision of the estimates can be determined based on the variance of the bootstrap estimate

$$\hat{\sigma}^2 = \frac{1}{B}\sum_{b=1}^{B} (\theta^{kb} - \bar{\theta}^{k*})^2.$$

6.5 Algorithm to bootstrap DEA

We have argued that the naive use of standard bootstrap methods is not satisfactory for DEA models, and we have discussed how to improve by smoothing and bias cor-

Table 6.6 Description of simplified version of `boot.sw98`

(1) Compute $\hat{\theta}^k$ as solutions to $\min\{\,\theta \mid (\theta x^k, y^k) \in \hat{T}\,\}$ for $k = 1, \ldots, n$.

(2) Use bootstrap via smooth sampling from $\hat{\theta}^1, \ldots, \hat{\theta}^K$ to obtain a bootstrap replica $\theta^{1*}, \ldots, \theta^{K*}$. This is done as follows

(2.1) Bootstrap, sample with replacement from $\hat{\theta}^1, \ldots, \hat{\theta}^K$, and call the results $\beta^1, \ldots, \beta^K$.

(2.2) Simulate standard normal independent random variables $\epsilon^1, \ldots \epsilon^K$.

(2.3) Calculate

$$\tilde{\theta}^k = \begin{cases} \beta^k + h\epsilon^k & \text{if } \beta^k + h\epsilon^k \le 1 \\ 2 - \beta^k - h\epsilon^k & \text{otherwise} \end{cases} \qquad \begin{array}{l}\text{(Smoothing and reflection}\\ \text{cf. page 172)}\end{array}$$

Note that by construction, $\tilde{\theta}^k \le 1$.

(2.4) Adjust $\tilde{\theta}^k$ to obtain parameters with asymptotically correct variance, and then estimate the variance $\hat{\sigma}^2 = \frac{1}{n}\sum_{k=1}^{K}(\hat{\theta}^k - \bar{\hat{\theta}})^2$ and calculate

$$\theta^{k*} = \bar{\beta} + \frac{1}{\sqrt{1 + h^2/\hat{\sigma}^2}}\,(\tilde{\theta}^k - \bar{\beta})$$

where $\bar{\beta} = \frac{1}{n}\sum_{k=1}^{K}\beta^k$.

(3) Calculate bootstrapped input based on bootstrap efficiency $x^{kb} = \frac{\hat{\theta}^k}{\theta^{k*}} x^k$.[1]

(4) Solve the DEA program to estimate θ^{kb} as

$$\theta^{kb} = \min\{\,\theta \ge 0 \mid y^k \le \sum_{j=1}^{K}\lambda_j y_j,\ \theta x^k \ge \sum_{j=1}^{K}\lambda_j x_j^{kb},$$
$$\lambda_j \ge 0,\ \sum_{j=1}^{K}\lambda_j = 1\,\} \qquad (k = 1, \ldots, n)$$

(5) Repeat the steps from (21) to obtain the bootstrap estimates

$$(\theta^{1b}, \ldots, \theta^{Kb}) \qquad (b = 1, \ldots, B)$$

(6) Calculate the mean and variance of $(\theta^{1b}, \ldots, \theta^{Kb})$ to get the bootstrap estimate θ^{k*}, the bias-corrected estimate $\tilde{\theta}^{k*}$, and the variance.

rection. We now present present a simplified method with smoothing of the method used in the R function `boot.sw98`.

It has been suggested that $B = 1000$ is suitable for calculating confidence intervals.

Table 6.7 Simplified version of `boot.sw98` in R

```
y <- cbind(1,2,3,4,5)
x <- cbind(2,4,3,5,6)

B <- 1000
thetaboot <- matrix(nrow=B, ncol=dim(x)[2])
thetati <- matrix(nrow=B, ncol=dim(x)[2])
# (1)
theta <- 1/dea(x,y,RTS=1,ORIENTATION=1)
N <- length(theta)
h <- 0.014  # bandwidth

# (2.1)
for ( b in 1:B) {
  beta <- sample(theta, N, replace=TRUE)
# (2.2)
  eps <- rnorm(N)
  thetatilde <- rep(0,N)
# (2.3)
  for (i in 1:N) {
    if ( beta[i]+h*eps[i] <= 1.0 ) {
      thetatilde[i] <- beta[i]+h*eps[i]
    } else {
      thetatilde[i] <- 2.0 -beta[i] -h*eps[i]
    }
  }
  thetati[b,] <- thetatilde
# (2.4)
  v = var(theta)
  thetastar = mean(beta) + (thetatilde-mean(beta))/(sqrt(1.+h^2/v))
# (3)
  xstar = theta/thetastar * x
  xstar = matrix(1,dim(x)[1],1) %*% theta/thetastar * x
# (4)
  thetaboot[b,] <-  1/dea(xstar,y,RTS=1,ORIENTATION=1)
} # for b
# done, now let's see the results
# (6)
print(colMeans(thetaboot),digits=3)
print(colMeans(thetati),digits=3)
bias <- colMeans(thetaboot) - colMeans(thetati)
print(bias,digits=3)
print(sd(thetaboot),digits=3)
boxplot(data.frame(thetaboot),boxwex=.5,ylim=c(min(thetaboot)-.1,1.05))
```

The DEA efficiency measures the radial distance in the input space from the observation point to the boundary of the technology set. We make a premature bootstrap of the efficiencies and use them to calculate the input vectors with this bootstrapped efficiency; this is done in step 3 in the above description. These bootstrapped input vectors are the inputs that determine the bootstrapped technology set in step 4 from which the final bootstrapped efficiency estimates are calculated. Note that x^{kb} is on the same ray as x^k. We could change this by also making the ray a random variable in the form of angles to be bootstrapped —i.e. by using polar coordinates to express x^k instead of the usual rectangular coordinates.

Please note that `boot.sw98` in FEAR bootstraps the Shaphard efficiency, and not Farrell efficiency as the R program does in Table 6.7. This is not a problem because the user has access to the individual bootstrap replica estimates in the component `boot` and then can just use `1/boot` for the Farrell bootstrap estimates.

6.5.1 Confidence intervals

As mentioned in Sect. 6.3.1, it is not advisable to calculate 95% confidence intervals because $\tilde{\theta}^k \pm 1.96\sigma_\theta$ as the distribution might not be a normal or symmetric; rather, it could be a skewed distribution or could have larger or smaller tails than the normal distribution. Instead, it is advisable to use the R function `quantile`. That is, to calculate a 95% confidence interval for firm 3, use

```
quantile(thetaboot[,3],probs=c(.025, .975),type=8)
```

If we do not include the firm index, here `3`, then the interval is based on all firms. This does not make any sense because the different firms have different efficiency levels, and we must determine the confidence interval for one firm at the time. For a 90% confidence interval, we just use `probs=c(.05,0.95)`. To determine the intervals for all firms, we can use

```
apply(thetaboot, 2, function(x) {
     quantile(x,probs=c(.025, .975), type=8, na.rm=TRUE) })
```

In the R function `boot.sw98` as part of the FEAR package, the confidence interval is estimated for the bias-corrected distance function values.

6.6 Numerical example in R

We will use the small examples from Table 6.8 to estimate the standard errors of the efficiency estimates and the confidence intervals for the input distance functions with a variable return technology. The R program including the data using the function `boot.sw98` is shown in Table 6.9.

The output is shown in Table 6.10. Note that if the aim is to obtain estimates of variance, the number of replicates, the value of the parameter `NREP`, must be at

Table 6.8 1 input og 1 output example

Firm	x	y
1	2	1
2	4	2
3	3	3
4	5	4
5	6	5

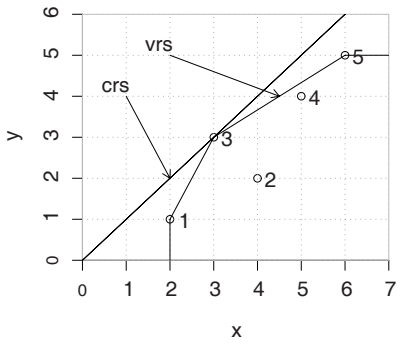


Table 6.9 Bootstrap DEA, R program

```
library(FEAR)
# Data
y <- cbind(1,2,3,4,5)
x <- cbind(2,4,3,5,6)

# DEA, Shephard input distance function,
d <- FEAR::dea(x,y, RTS=1, ORIENTATION=1)
# Efficiencies
print(1/d,digits=3)
print(mean(1/d),digits=3)

#  Bootstrap
b <- boot.sw98(x,y, RTS=1, ORIENTATION=1, NREP=2000)
print(b,digits=3)
print(sqrt(b$var),digits=3)
```

least 50; correspondingly, to obtain confidence intervals, at least 100 are required. It might also be necessary for the number of replicates to be much larger to obtain stable results for larger datasets; however, that relation has not been tested as of this writing. Part of the output is the individual replications, returned as item `boot`. All of the output items are described in the help file for `boot.sw98` in the FEAR package; from inside R, we use the command `?boot.sw98`. In the last line, we have calculated the standard error of the input distance, the square root of the variance.

The above method is very simple to use in practice. However, it does have a pedagogical drawback: everything is hidden in the function `boot.sw98`. To make up for this, we mimicked the function in R statements to see the inner working of bootstrap in DEA, just as we did for the traditional bootstrap procedure in section 6.3 on page 165.

The bias-corrected estimate is in item `dhat.bc` and can also be found by subtracting the bias from the DEA estimate of the distance function value, item `dhat`; i.e. `b$dhat − b$bias`. The confidence interval is estimated around the bias-

Table 6.10 Output from bootstrap

```
> # Efficiens
> print(1/d,digits=3)
[1] 1.000 0.625 1.000 0.900 1.000
> print(mean(1/d),digits=3)
[1] 0.905
> #  Bootstrap
> print(b,digits=3)
$bias
[1] -0.143 -0.151 -0.130 -0.101 -0.150

$var
[1] 0.00914 0.01061 0.00707 0.00538 0.01358

$conf.int
     [,1] [,2]
[1,] 1.01 1.35
[2,] 1.61 1.99
[3,] 1.01 1.30
[4,] 1.12 1.42
[5,] 1.00 1.41

$dhat
[1] 1.00 1.60 1.00 1.11 1.00

$dhat.bc
[1] 1.14 1.75 1.13 1.21 1.15

$boot
      [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]  [,8]  ...
[1,] 0.560 0.565 0.565 0.572 0.572 0.576 0.590 0.591  ...
[2,] 0.983 0.986 1.027 1.037 1.086 1.092 1.107 1.119  ...
[3,] 0.569 0.602 0.611 0.614 0.622 0.626 0.628 0.631  ...
[4,] 0.662 0.701 0.702 0.708 0.713 0.713 0.723 0.727  ...
[5,] 0.533 0.537 0.541 0.543 0.544 0.545 0.550 0.551  ...
...
     [,1997] [,1998] [,1999] [,2000]
[1,]    1.00    1.00    1.00    1.00
[2,]    1.60    1.60    1.60    1.60
[3,]    1.00    1.00    1.00    1.00
[4,]    1.11    1.11    1.11    1.11
[5,]    1.00    1.00    1.00    1.00

> print(sqrt(b$var),digits=2)
[1] 0.096 0.103 0.084 0.073 0.117
```

corrected estimate. The default confidence interval is 95% but can be changed using the option `alpha`. Either a scalar option or a vector option is available, indicating the statistical sizes of the confidence intervals to be estimated. Thus, `alpha=.1` will calculate limits corresponding to a $1.0 - 0.1 = 90\%$ interval.

To explain the confidence interval further, let us recalculate the Shephard input values to Farrell input values by calculating the reciprocal. This is done below where the output from the R commands is also shown.

```
> 1/b$dhat
[1] 1.000000 0.625000 1.000000 0.900009 1.000000
> 1/b$dhat.bc
[1] 0.8764797 0.5707759 0.8855686 0.8228137 0.8705459
> 1/b$conf.int[,c(2,1)]
          [,1]      [,2]
[1,] 0.7439961 0.9932824
[2,] 0.5030548 0.6218341
[3,] 0.7764515 0.9935884
[4,] 0.7085692 0.8951720
[5,] 0.7082100 0.9940264
```

Because of the reciprocal property, the upper limit becomes the lower limit and vice versa, and that is why the index in `$conf.int` is reversed. These numbers indicate that the upper limit of the confidence interval `1/b$conf.int` is very close to the estimated efficiency `1/b$dhat`, whereas the lower limit is far below. The closeness of the upper limits and the efficiencies means that the frontier corresponding to the upper limit coincides with the DEA-estimated frontier. The lower limit in the confidence interval for the efficiencies corresponds to a frontier to the left of the DEA frontier; if we measure the efficiency of the observations against this frontier, we get the lower limits of efficiency; this frontier is shown in Figure 6.5 on the next page as a dotted frontier. This frontier corresponding to the lower limit of the efficiencies is far from the efficiency estimates because a variation in inputs during the bootstrap procedure in which the input gets smaller will enlarge the technology set and move it to the left (as the new input can be outside the frontier) and will therefore create a new frontier. A larger input, on the other hand, will mostly leave the frontier unchanged because it will be below the already existing frontier. Note that bias-corrected efficiency is more likely to be in the middle of the confidence interval because bias correction is intended to correct for the derived bias or skewness in the DEA estimation.

6.7 Interpretation of the bootstrap results

To further example how to interpret the DEA bootstrap results, let us investigate two special cases. The first contains just one input and one output, whereas the second contains two inputs and one output.

6.7.1 One input, one output

Let us take a closer look at the output in Table 6.10 on page 178 from the R commands in Table 6.9. This is a small problem involving 5 firms, 1 input, and 1 output. Bootstrap is conducted using the method `boot.sw98`, and the output includes several components or items. The item named `$dhat` is the estimated Shephard input distance Function, which is equal to the reciprocal of technical input efficiency, Farrell efficiency; i.e. $TE = \frac{1}{\texttt{b\$dhat}}$. We can see this by comparing lines 8, 13, and 34. The bias-corrected Shephard input distance function is found to be `$dhat.bc` in line 37. The bias-corrected Shephard input distance functions can also be found by subtracting the bias from the DEA estimates; i.e. as `$dhat` − `$bias` ; cf. our discussion of this idea in Sect. 6.4.3 on page 173.

If the bias-estimated distance input function value is $\tilde{\theta}$, then a point on the bias-corrected frontier is $\frac{1}{\tilde{\theta}}x$ where x is the observation of the input. Because we are looking at input functions and input efficiency, the output y remain the same.

We can plot the observations and the input corresponding to the bias-corrected Sheppard input distance function by

```
dea.plot.frontier(x,y,txt=1:N)
dea.plot.frontier(x/b$dhat.bc,y,lty="dashed",add=T)
dea.plot.frontier(x/b$conf.int[,2],y,lty="dotted",add=T)
```

The options `lty` specify the line type; the default is `solid`.

The resulting figure is shown in Fig. 6.5. If we were to draw a random sample

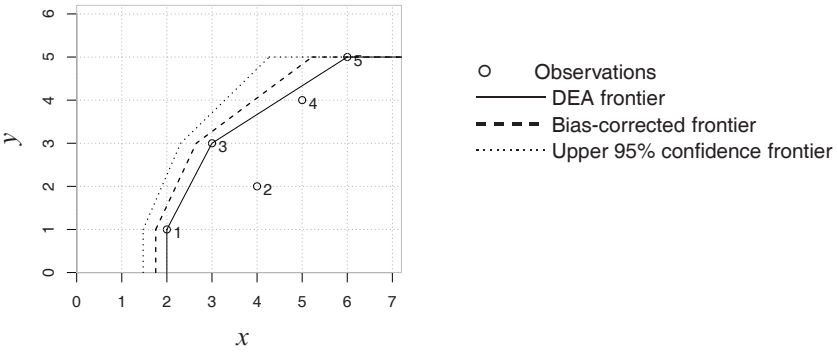


Fig. 6.5 Bias-corrected frontier, input direction

to estimate the frontier, it would to the right of the 95% confidence frontier with a probability of 95%.

Another way to demonstrate efficiency and confidence intervals is As shown in Fig. 6.6, constructed using the following R commands:

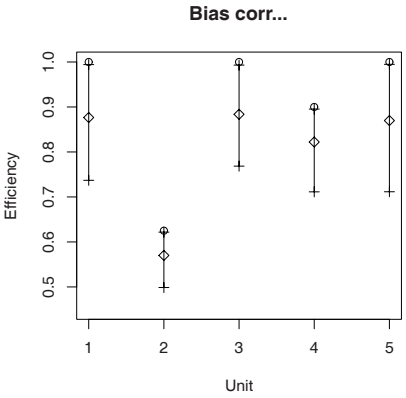


Fig. 6.6 Bias-corrected efficiency estimates (◇), DEA estimates (○) and 95% confidence limits—one input, one output

```
plot(1/b$dhat,ylim=c(.45,1),main="Bias corr...",
     xlab="Firm", ylab="Efficiency")
points(1/b$dhat.bc,pch=5)
for ( i in 1:5 )lines(rep(i,2),1/b$conf.int[i,],type="o",pch=3)
```

6.7.2 Two inputs

The isoquants for the two inputs are calculated using the following R program, which is similar to the program for one input and one output in Table 6.9 on page 177 except that the isoquant is plotted instead of the frontier. To plot the isoquant, we have normalized the inputs with the output and then used an output of 1 for all firms because then all firms have the same isoquant and can be compared. Thus, implicitly, we are assuming constant returns to scale.

```
# The data
y <- t(matrix(c(1,2,3,1,2)))
x <- t(matrix(c(2,2,6,3,6,  5,4,6,2,2), ncol=2))
N <- dim(x)[2]
x1 = x[1,]/y
x2=x[2,]/y
# The frontier for the technologies
dea.plot.isoquant(x1,x2,txt=1:N)
# The observations have dotted lines from origo
for ( i in 1:length(y) ) {
   lines(c(0,x1[i]), c(0,x2[i]),lty="dotted")
}
# bootstrap
b <-boot.sw98(rbind(x1,x2),matrix(rep(1,N),nrow=1),NREP=2000,RTS=3)
dea.plot.isoquant(x1/b$dhat.bc,x2/b$dhat.bc,lty="dashed",add=T)
```

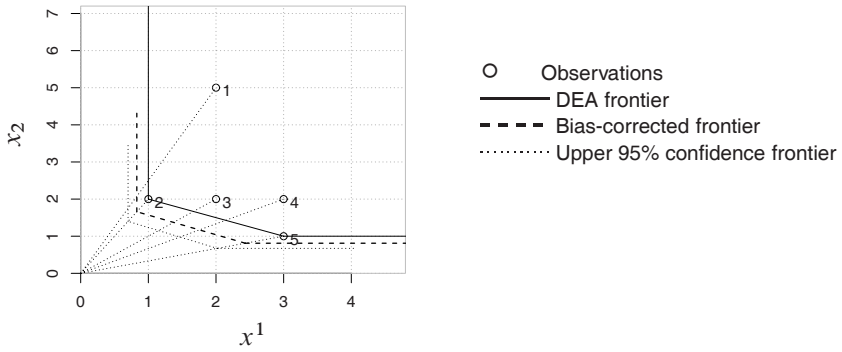


Fig. 6.7 Bias-corrected frontier, input direction, 2 inputs

```
dea.plot.isoquant(x1/b$conf.int[,2],x2/b$conf.int[,2],lty="dotted",add=T)
```

The graphs are in Fig. 6.7. Again, we can see that the bias-corrected frontier is below the Isoquant, making the technology set larger, and that the upper confidence limit is increasing it even further.

The graph in Fig. 6.8 is made using the R program lines

```
plot(b$dhat,ylim=c(1,3),main="Bias␣corr...",
       xlab="Firm",ylab="Distance␣function")
points(b$dhat.bc,pch=5)
for ( i in 1:5 )lines(rep(i,2),b$conf.int[i,],type="o",pch=3)
```

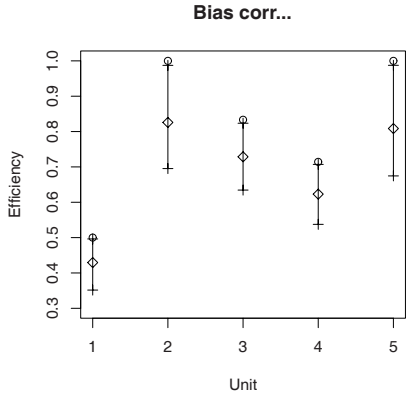


Fig. 6.8 Bias-corrected efficiency estimates (◇), DEA estimates (○) and 95% confidence limits – two inputs

6.8 Statistical tests using bootstrapping

Let us finally illustrate how to use bootstrapping to test hypotheses. Specifically, we will show how to tests a returns to scale hypothesis, but other tests can be developed along the same lines.

We wish to test whether the technology set T from which our observations are sampled exhibits constant returns to scale. Formally, we wish to test the hypothesis that the technology exhibits constant returns to scale against the alternative, that it is VRS:

$$H_0 : T \text{ is CRS}$$
$$H_A : T \text{ is VRS}$$

If we reject H_0, then we can test if the technology set is DRS, but we will leave that project to the reader.

If the hypothesis is true, then the efficiencies calculated from the VRS technology are the same as the efficiencies calculated from the CTS technology. If there is not CRS, then at least one of the efficiencies will be different; i.e. CRS efficiency will be smaller than VRS efficiency. One way to examine this is to see whether the scale efficiency, cf. page 99,

$$SE^k = \frac{E^k_{\text{CRS}}}{E^k_{\text{VRS}}} \qquad (k = 1, \ldots, K)$$

is equal to 1 for all firms, meaning that the technology is CRS, or whether there is at least one firm where it is less than 1, meaning that the technology is VRS. For a given set of observations of K firms, we must therefore reject the hypothesis if at least one of the estimated SE has a value less than 1. However, as the connection between the technology set and the scale efficiencies is an uncertain or stochastic connection, we must reject the hypothesis if at least one of the estimated SE has a value significantly less than 1, i.e. if one of the estimated SE is less than a critical value. The problem is then to compute this critical value.

Instead of looking at the scale efficiencies individually, we could look at the test statistic

$$S^1 = \frac{1}{K} \sum_{k=1}^{K} \frac{E^k_{\text{CRS}}}{E^k_{\text{VRS}}}$$

or the one that we are going to use in the following:

$$S = \frac{\sum_{k=1}^{K} E^k_{\text{CRS}}}{\sum_{k=1}^{K} E^k_{\text{VRS}}}. \tag{6.1}$$

If the H_0 is true, then S will be close to 1, and if the alternative is true, then $S < 1$. As $S \leq 1$ by construction, we will reject H_0 if S is significantly smaller than 1. We therefore seek a critical threshold for the statistic S; if it is smaller than this

value, then we will reject the hypothesis. Thus, we seek a critical value c_α that will determine whether we reject H_0, the hypothesis of constant returns to scale, if $S < c_\alpha$ and $\Pr(S < c_\alpha \mid H_0) = \alpha$ where α is the size of the test, typically 5% ($\alpha = 0.05$). The size of the test, α is the probability of rejecting the hypothesis even though it is true. (This is a type I error.)

Unfortunately, we do not know the distribution of S under H_0, and therefore, we cannot calculate c_α directly. One way to address this lack of distributional knowledge is to use a bootstrap method, and we will now show that one can bootstrap the distribution of S under H_0. We show how this can be done using a very small example: the data from Table 6.8 on page 177. First, we enter the data and calculate the statistic S and its quantile using the following commands in R:

```
library(FEAR)
y <- cbind(1,2,3,4,5)
x <- cbind(2,4,3,5,6)
e <- 1/dea(x,y,RTS=3)
ev <- 1/dea(x,y,RTS=1)
sum(e)/sum(ev)
nrep <- 2000
Bc <- boot.sw98(x,y,NREP=nrep,RTS=3)
Bv <- boot.sw98(x,y,NREP=nrep,RTS=1,XREF=x,YREF=y,DREF=1/e)
s <- colSums(1/Bc$boot)/colSums(1/Bv$boot)
quantile(s,c(1,2,5,10,15,30,50)/100.0)
```

We calculate the CRS efficiency (`RTS=3`), the VRS efficiency (`RTS=1`), and the test statistic S from (6.1). The following lines is the bootstrap. First, the variable `nrep` is set to the number of bootstrap replications that we will use. Then, we bootstrap under the null-hypothesis. Thereafter, we bootstrap under the alternative while assuming that H_0 is in fact true by using the option `DREF=1/e` where `1/e` is efficiency calculated under the CRS technology.

The output is shown in Table 6.11. The estimate of S is 0.802945, which seems to

Table 6.11 Output for test of constant returns to scale

```
> y <- cbind(1,2,3,4,5)
> x <- cbind(2,4,3,5,6)
> nrep <- 2000
> e <- 1/dea(x,y,RTS=3)
> ev <- 1/dea(x,y,RTS=1)
> sum(e)/sum(ev)
[1] 0.802945
> Bc <- boot.sw98(x,y,NREP=nrep,RTS=3)
> Bv <- boot.sw98(x,y,NREP=nrep,RTS=1,XREF=x,YREF=y,DREF=1/e)
> s <- colSums(1/Bc$boot)/colSums(1/Bv$boot)
> quantile(s,c(1,2,5,10,15,30,50)/100.0)
       1%        2%        5%       10%       15%
30%       50%
0.7409859 0.7431850 0.7472870 0.7531393 0.7585869 0.7940538 0.8561436
```

be far less than 1, but we only have 5 firms, and the output from `quantile` shows that .80 corresponds to a little more than 30%. Therefore, there is a 30% probability of observing a lower value of S than the one we obtained, and therefore, we do not reject H_0; i.e. we do not reject that there exist constant returns to scale. If we were to make further calculations under this model, we would therefore assume constant returns to scale and use a CRS technology.

Earlier, we introduced the idea of the critical value, which can be calculated using the function `critValue`, which takes the bootstrapped statistics and the size of the test as input. We also have at our disposal the function `typeIerror`, which calculates the probability of type I error: the probability of rejecting the hypothesis if it is true.

```
critValue <- function(s,alfa)  {
  ss <- sort(s)
  mean( ss[floor(alfa*length(s))], ss[ceiling(alfa*length(s))] )
}

typeIerror <- function(shat,s) {
  reject <- function(alfa)  {
    quantile(s,alfa,names=F) - shat
  }
  uniroot(reject,c(0,1))$root
}
```

Both functions are part of the Benchmarking package. Using the two functions with the data above yields the output

```
> shat <- sum(e)/sum(ev)
> shat
[1] 0.802945
> critValue(s,0.05)
[1] 0.7418619
> typeIerror(shat,s)
[1] 0.3337649
```

Thus, if the estimated value of S is less than the critical value 0.7418619, we reject the hypothesis. Correspondingly, because the estimate of S, `shat`, is 0.802945, we do not reject the hypothesis. The results obtained using `typeIerro` show that there is a probability of 0.3337649 that one will obtain a lower estimate of S than the one we found, or in other words, that we will be making a mistake if we reject the hypothesis on the basis of our estimate.

6.9 Summary

DEA originates in the operations research and management science, and this means that the evaluation of DEA models is not a purely statistical exercise. Indeed, historically the use of traditional statistical tests has not been emphasized. Considerable progress has however been made in this respect over the last 15 years, and we introduced some important contributions in this chapter.

One possibility is to use general non-parametric tests, i.e. tests used when the underlying distribution is unknown, like Kolmogorov–Smirnov tests and Kruskal–Wallis tests. Such tests can be used to evaluate a series of different assumptions and hypothesis but as always they may suffer from limited power.

Another possibility is to rely on parametric tests. If we can make reasonable assumptions regarding the underlying distribution of inefficiency and noise in the data, a series of tests are possible. We discussed tests for group differences and tests for model assumptions. To justify the distributional assumptions in a parametric approach, we may rely on asymptotic theory, i.e. theoretical properties that can only be established for large samples. Simulation studies based on samples of moderate size suggests that such assumptions may well be justified in many applications.

A third approach, and one that has become particularly popular with the development of effective computer programs, is the use bootstrapping. The bootstrap is a computer-based method that can answer many statistical questions. The approach replicates sampling uncertainty by creating repeated samples of the original sample. We spend most of this chapter covering bootstrap-based inference in DEA models. In particular, we showed how to make bias corrections and construct bias corrected confidence intervals for the individual efficiencies. One advantage of R is that effective bootstrapping methods for DEA models have been made easily available, not the least via the FEAR package.

In the appendix, we discuss the use of statistical methods in second-stage analyses, i.e. analyses performed after the development of a benchmarking model, to validate the model and to explore the possible causes of the variations in efficiencies. A common approach in such studies is tobit regression, and we discuss how to perform and interpret such an analyses.

6.10 Bibliographic notes

Consistency of DEA estimates and asymptotic tests are based on Banker (1993) and Banker (1996).

The bootstrap method was invented in 1979 and it is now a well established statistical method. A good reference to the statistical theory of bootstrap with lot of examples is Efron and Tibshirani (1993); the mathematical level of the book is moderate. Our description of the bootstrap, and in particular Table 6.1 is taken from that book. A more advanced text assuming a grounding in statistics is Davison and Hinkley (1997). The reflection method is described in (Silverman, 1986, 30).

R is based on S, a language and an environment for data analysis. Bbootstrap methods have been in S almost since the beginning (Chambers and Hastie, 1992).

Bootstrap of DEA model have a winding history, the first attempt was done around 1992. The bootstrap method for DEA described in this book is from Simar and Wilson (1998) and Simar and Wilson (2000). Their approch is implemented in R as `boot.sw98` as part of the FEAR library (Wilson, 2008). The simplified

description of `boot.sw98` in Table 6.6 by and large follows Simar and Wilson (1998).

The tobit model covered in the Appendix was first used by Tobin in 1958 (Tobin, 1958), is discussed in many textbooks, including Greene (2008) and Maddala (1983). The tobit model is traditionally used with point of truncation at 0, which makes the marginal impact relatively easy to calculate. Because efficiency scores are truncated at 1, we have derived the marginal impact for this case. An important critical paper on the tobit approach in benchmarking, Simar and Wilson (2007), instead proposes the use of bootstrapping. Hoff (2007) also identifies a number of theoretical issues associated with current practice, but she concludes after analyzing an actual dataset that the tobit procedure does produce reasonable estimates and, moreover, can be substituted for by a regular OLS approach under some conditions. McDonald (2009) questions whether the DEA scores should be seen as a censored distribution, arguing for the use of a "fractional" model, but he also concludes that theoretical niceties are of little concern to "instrumentalists", and that hundreds of two-stage DEA studies have proven very useful in providing insight into real-world production processes.

6.11 Appendix: Second stage analysis

When we have estimated the efficiencies of the firms in an industry, we often become interested in understanding why some firms are more efficient than others. Is their efficiency related to firm size, CEO age, the fraction of highly educated employees at the firm, the use of ICT, the business environment in different regions, and/or other factors?

We may also wonder if the variations in estimated efficiency really reflect variations in performance or if we may have left out important inputs or output (i.e. we might be interested in validating the model). Should we have included a measure of soil quality in a farming model, a measure of socio-economic status of the model examining students in a school, or a measure of quality in a hospital model? In developing a benchmarking model for German DSO regulation, cf. Sect. 10.3, we did, for example, make a final evaluation of several hundreds of omitted candidate variables.

Both aims are often pursued using what is commonly called second-stage analysis, i.e. post-efficiency analysis that aims to explain the variations and validate the model. In this appendix, we discuss the use of statistical methods in second-stage analyses. The relevance of such analyses and the corresponding methods is not restricted to DEA studies. Other best-practice results can be analyzed using the same methods.

To investigate if *categorial variables* like high/low, east/west, and low/medium/high may explain some of the variation, we can use a number of non-parametric tests: e.g. the Mann-Whitney-Wilcoxon rank-sum test. This is a non-parametric test used to assess whether two independent samples of observations have equally

large values. This process is largely equivalent to performing an ordinary parametric two-sample *t*-test on the data after ranking the combined sample. We can also use other non-parametric tests like the Kolmogorov-Smirnov and Kruskal-Wallis tests, as demonstrated in Sect. 6.2.1. All tests can easily be undertaken in R.

The most common approach used to investigate if a set of *continues variables* variables may explain the variations in efficiency is to conduct a tobit regression. Tobit regression is similar to ordinary regression analysis except that the noise term is truncated. The use of this method in a benchmarking context is the focus of some debate in the literature (cf. below), but it is widely applied and is generally considered to be useful.

Let E be the Farrell input efficiency calculated in a DEA model, an SFA model or some combination of models (cf. e.g. the combined use of several models in regulatory benchmarking as explained in Chap. 10). We will return to models of output efficiency later. We are now interested in modeling how E depends on other variables $z = (z_1, z_2, \ldots, z_q)$. That, is we would like to estimate a model

$$E = g(z, a).$$

whereby efficiency E is explained by the variables z and parameters a.

6.11.1 Ordinary linear regressions OLS

A model is a linear regression ,model

$$E = a_1 1 + a_2 z_2 + \cdots + a_q z_q + \varepsilon = az + \varepsilon$$

where ε is a random error that reflects that the model does not completely explain the efficiency levels. It is easy to estimate this model using OLS. In R, this can be done using the function `lm`.

One advantage of this approach is that it is easy to find the marginal effect on efficiency based on a marginal change in z_j:

$$\frac{\partial E}{\partial z_j} = a_j,$$

Because this effect is independent of the value of all the variables, it is also easy to interpret—it shows how much the efficiency tends to increase if a_j is increased by one unit.

Although ordinary regressions are widely used in practice, they suffer from a theoretical problem in a benchmarking setting. They do not take into account that efficiencies are greater than 0 and less than or equal to 1 and that many efficiencies are typically at the upper boundary of 1. There is nothing in the method that ensures that the fitted value, the expected value, or the mean will be less than or equal to 1. The tobit model for censored regression can be used to solve this problem.

6.11.2 Tobit regression

When the dependent variable is censored, we do not observe the underlying values of this variable in all cases. Values in a specific range are reported as a single value. In the case of E, we can see the underlying efficiencies as a stochastic variable and the observation of efficiency E as a a censored version hereof where values below 0 are reported as 0 and values above 1 are reported as one. Therefore, the model becomes

$$E = \begin{cases} 0, & \text{if } az + \varepsilon \leq 0 \\ az + \varepsilon & \text{if } 0 < az + \varepsilon < 1 \\ 1 & \text{if } az + \varepsilon \geq 1 \end{cases}$$

Our challenge is to estimate a on the basis of the observed efficiencies E^k from K firms $k = 1, \ldots, K$.

In general, we do not have any firms with reported efficiency of 0. Therefore, let K_1 be the number of firms for which $E = 1$ (i.e. the number of efficient firms) and K_0 be the number of firms for which $E < 1$. We then have $K = K_0 + K_1$.

The probability that $E = 1$ is the probability that $az + \varepsilon >= 1$. Let F be the probability distribution function for ε and f the corresponding density function. Then the probability of $E = 1$ is

$$\begin{aligned} \Pr(E = 1) = \Pr(az + \varepsilon \geq 1) &= 1 - \Pr(az + \varepsilon < 1) \\ &= 1 - \Pr(\varepsilon < 1 - az) = 1 - F(1 - az), \end{aligned}$$

and the probability that $E = 0$ is

$$\Pr(E = 0) = \Pr(az + \varepsilon \leq 0) = \Pr(\varepsilon < -az) = F(-az).$$

The case where in which $0 < E < 1$ corresponds to $E = az + \varepsilon$ or $\varepsilon = E - az$ such that the density is this case [ED21]is $f(E - az)$.

The likelihood function for K observations of efficiencies is then given as the product of the K individual terms for the cases mentioned above.

$$\begin{aligned} L &= \prod_{k:E^k=1} \Pr(E^k = 1) \prod_{k:0<E^k<1} f(E^k - az^k) \\ &= \prod_{k:E^k=1} \left(1 - F(1 - az^k)\right) \prod_{k:0<E^k<1} f(E^k - az^k). \end{aligned}$$

We have here not taken into account that E in the theory could be equal to 0. Because the number of such observations is 0, the corresponding likelihood factor is 1 irrespective of the value of $\Pr(E = 0)$.

To estimate the above model, we also need to choose a probability distribution F. The most commonly used distribution is the normal distribution, and in this case, the model is called the *tobit regression* model. We will not formulate the likelihood

function in this particular case but will instead refer the reader to the literature mentioned in the bibliographic notes. The actual optimization process is conducted using standard iterative optimization routines that are also available in R. As part of the estimation process using standard programs, the variance of the estimated parameters is also calculated such that statistical inference is possible.

Now, in benchmarking applications, we are typically interested in knowing the marginal effect of a marginal change in one of the explanatory variables z. In the OLS framework, these effects are readily available as the parameter estimates a. In the tobit framework, they are more difficult to determine, and we will provide them here.

In the rest of this section, we use EV for the mean or expectation of a random variable to be able to distinguish the mean EV from efficiency E. We are interested in knowing how $\mathrm{EV}(E|z)$ varies with z, i.e. how a change in z influences efficiency E on average. The conditional expectation consists of three parts corresponding to the three parts of the model for E.

$$\begin{aligned}\mathrm{EV}(E|z) &= \int E\, d\Pr(E|z) \\ &= \int 0\, d\Pr(E=0|z) + \int E\, d\Pr(0<E<1|z) + \int 1\, d\Pr(E=1|z) \\ &= \int_{-az}^{1-az} \varepsilon\, d\Pr(\varepsilon|z) + 1 - \Pr(\varepsilon < 1-az|z).\end{aligned}$$

where there last equality can be verified by inserting the definition of E and making a few reformulations.

We now calculate the two probability terms separately. The last is simple to calculate when we assume that the error term is normally distributed, i.e. $\varepsilon \sim N(0,\sigma^2)$. The first term is slightly more complicated because it involves real integration. The final result is that

$$\begin{aligned}\mathrm{EV}(E|z) = az\left(\Phi\Big(\frac{1-az}{\sigma}\Big) - \Phi\Big(\frac{-az}{\sigma}\Big)\right) \\ + \sigma\left(\varphi\Big(\frac{-az}{\sigma}\Big) - \varphi\Big(\frac{1-az}{\sigma}\Big)\right) + 1 - \Phi\Big(\frac{1-az}{\sigma}\Big).\end{aligned}$$

Although this process looks complicated, the terms can interpreted simply based on the defining equation. The last two terms, $1-\Phi$, correspond to the effect of the firms where $E=1$ multiplied by the probability of this event. The first term is the linear effect az multiplied by the probability that $0<E<1$. The second term is the effect of the error term ε. In the linear model, the OLS model, this effect is zero because the expected value of ε is 0, but here, the mean of ε is conditioned to the interval where $0<az+\varepsilon<1$, i.e. $-az<\varepsilon<1-az$.

Based on the above, we can also find

$$\mathrm{EV}(E|0 < E < 1, z) = az + \sigma \frac{\varphi\left(\frac{-az}{\sigma}\right) - \varphi\left(\frac{1-az}{\sigma}\right)}{\Phi\left(\frac{1-az}{\sigma}\right) - \Phi\left(\frac{-az}{\sigma}\right)} = az + \sigma\, M(az)$$

where the function $M(\cdot)$ is called the *inverse Mills ratio*.

Now we can determine how $\mathrm{EV}(E|z)$ varies with z by finding the derivative of $\mathrm{EV}(E|z)$ w.r.t. z. To do so, we must find the derivatives of the individual terms in $\mathrm{EV}(E|z)$. We will not present the details here, but we should note that they make use of the chain rule and the fact that Φ is the antiderivative of ϕ such that $\Phi' = \varphi$ and $\Phi(t) = \int_{-\infty}^{t} \varphi(\varepsilon)\, d\varepsilon$. By collecting terms and canceling out where possible, we get

$$\frac{\partial\, \mathrm{EV}(E|z)}{\partial z_h} = a_h \left(\Phi\left(\frac{1-az}{\sigma}\right) - \Phi\left(\frac{-az}{\sigma}\right) \right). \tag{6.2}$$

Again, the results are easy to interpret: the term a_h corresponds to the linear term that we also found for the OLS model in Sect. 6.11.1, but here, it is corrected for the probability that $0 < E < 1$. If $E = 0$ or $E = 1$, then a marginal change in z will not change E.

All of the above calculations can be easily done numerically; both Φ and φ are available as functions in R, as we shall see in the numerical example.

Output efficiency and tobit

For output efficiency F, we have $F \geq 1$; therefore, the model is

$$F = \begin{cases} az + \varepsilon & \text{for } az + \varepsilon > 1, \\ 1 & \text{otherwise,} \end{cases}$$

where there is no upper bound; the bound that was an upper bound for input efficiency E is here a lower bound. To determine the expectation of F, we use some of the same terms as before. However, we use them a little differently and derive

$$\mathrm{EV}(F|z) = \Phi\left(\frac{1-az}{\sigma}\right) + az\left(1 - \Phi\left(\frac{1-az}{\sigma}\right)\right) + \sigma\, \varphi\left(\frac{1-az}{\sigma}\right),$$

and the derivative w.r.t. z_h becomes

$$\frac{\partial\, \mathrm{EV}(E|z)}{\partial z_h} = a_h \left(1 - \Phi\left(\frac{1-az}{\sigma}\right)\right).$$

Again, this corresponds to the derivative of the expected figure for input efficiency where the upper bound is now the lower bound and the upper bound is infinity. The interpretation is also as before; the linear effect a_h is multiplied by the probability that $F > 1$, i.e. 1 minus the probability that $F = 1$.

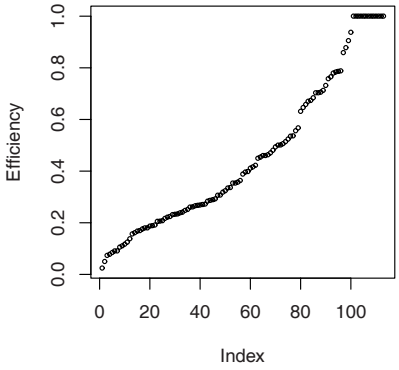


Fig. 6.9 Efficiency in Norwegian forestry

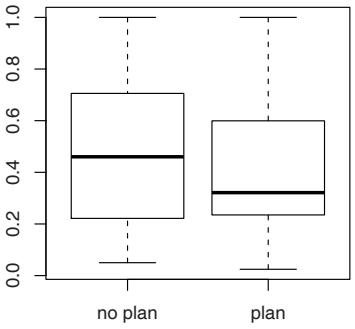


Fig. 6.10 Explaining efficiency by the absence or presence of a forest plan ($z_6 = \mathbf{0}, \mathbf{1}$)

6.11.3 Numerical example in R

We use a data set for 113 farmers in forestry in Norway. The basic DEA model is quite simple; it includes just two inputs and one output. The input variables are the value of the woodland and variable cost, and the output is earned profit. The variables that we will later use to explain efficiency, are secondary income from ordinary farming (z_1), owner age (z_3), and whether there is a long-term plan (z_6).

The input efficiencies in a variable-returns-to-scale DEA technology are shown in sorted order in Fig. 6.9. We see that there is tremendous variation in efficiency levels and that only a few firms are fully efficient. We may therefore ask what might explain this variation and what additional variables we should perhaps have included in the DEA model.

The efficiencies were calculated using the R script in Table 6.12 on the facing page, where we have also included the second step: an OLS regression and a tobit regression. The function `tobit` used to conduct tobit regressions is part of the AER package. The tobit regression is the R method `tobit` called with an input formula just like `lm` for linear regression. Numerical differences may affect the convergence, and we therefore ended up rescaling the z_1 variable by dividing the original values by 10^6; this process yielded a maximal value of 2,49.

In Fig. 6.10, the empirical box plot indicates that firms without a plan are more efficient than firms with a plan. However, the tendency is only vague, and in the OLS regression, the parameter for the plan factor, z_6 is estimated at $-.016$, which indicates that a firm with a forest plan has an efficiency level that is 1.6 percentage points lower. The standard error of the estimate is relative large, and the t-value of 0.76 shows that the parameter is not at all significantly different from zero. The same

Table 6.12 Two-stage DEA in R

```
> library(Benchmarking)
> library(AER)
> d <- read.csv("norWood2004.csv", header=T, comment.char = "#")
> x <- cbind(d$x,d$m)
> y <- d$y
> e <- dea(x,y)
> E <- eff(e)
> eOls <- lm(E ~ z1+z3+z6, data=d)
> summary(eOls)

Call:
lm(formula = E ~ z1 + z3 + z6, data = d)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.850e-01  1.503e-01   1.231   0.2210
z1          -1.023e-07  6.062e-08  -1.688   0.0943 .
z3           7.425e-03  2.962e-03   2.507   0.0137 *
z6          -1.635e-02  5.479e-02  -0.298   0.7659
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> zz1 <- d$z1/1e6
> eTob <- tobit(E ~ zz1+z3+z6, left=-Inf, right=1, data=d)
> summary(eTob)

Call:
tobit(formula = E ~ zz1 + z3 + z6, left = -Inf, right = 1, data = d)

Observations:
         Total  Left-censored     Uncensored Right-censored
           113              0            100             13

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.165955   0.165135   1.005   0.3149
zz1         -0.125615   0.066745  -1.882   0.0598 .
z3           0.008456   0.003265   2.590   0.0096 **
z6          -0.010403   0.060475  -0.172   0.8634
Log(scale)  -1.171818   0.073290 -15.989   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

estimated parameter in the tobit model is -0.01 with a t-value of -0.17 that is also not significantly different from zero. Therefore, the tendency we see in the numbers is probably purely incidental; it is likely that having a plan does not influence efficiency.

In Fig. 6.11, the efficiencies are plotted against the variable z_1, secondary income from ordinary farming. The tendency in the figure is that the larger the secondary

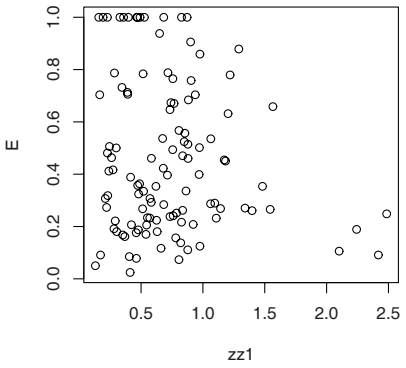


Fig. 6.11 Explaining efficiency by secondary income z_1 (rescaled to `zz1`)

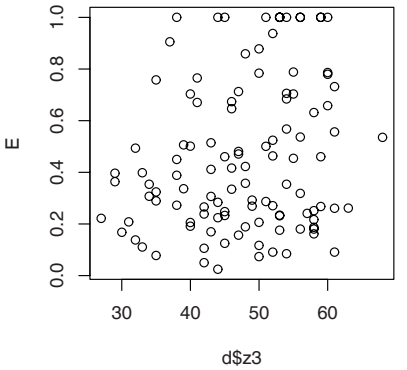


Fig. 6.12 Explaining efficiency by the age of owner z_3

Table 6.13 Tobit model with continuous age and age below 37 as an explanation

Model		Intercept	z_1	Age	z_6
Age continuous	Estimate	0.166	-0.126	0.008	-0.010
	z value	1.0	-1.9	2.6	-0.2
Age < 37	Estimate	0.594	-0.127	-0.202	-0.000
	z value	9.8	-1.9	-2.3	-0.0

income, the lower the efficiency level. This may be because farmers spend more time on secondary work and therefore neglect wood farming to some degree, which will lead to lower efficiency. The estimated parameter in the OLS regression for variable z1, determined using method `lm` as indicated in Table 6.12, is negative. This supports the impression, based on the figure, that higher secondary income is associated with lower efficiency. The parameter is only significantly different from zero at a 10% level; the t-value is only 1.77.

In Fig. 6.12, the age of the owner z_3 is plotted against efficiency, and it emerges that the effect of age is positive and significantly different from zero. The older the owner, the more efficient the firm. This may indicate that forestry farming is learned during the practice of forestry. From the figure, we can see that the increase only occurs below the age of 37. Instead of using age `z3` as a continuous variable, we can also use it as a factor with levels under 37 and over 37. The command used to estimate a tobit model, where age is this two-level factor, is

```
tobit(E~zz1+as.factor(d$z3<37)+z6, left=-Inf, right=1, data=d)
```

and the results are shown in Table 6.13, where the estimates achieved using age z_3 as a continuous variable are also shown. The difference between the two tobit models

Table 6.14 Comparing marginal effects in the Norwegian forest model

	$z_1 \cdot 10^{-6}$	z_3	z_6
OLS	-0.102	0.00742	-0.01635
Mean of effect for all firms	-0.114	0.00769	-0.00946
Effect at mean value of z	-0.116	0.00782	-0.00962
Effect at min	-0.100	0.00673	-0.00828
Effect at max	-0.111	0.00748	-0.00920

is minimal. The conclusion is that age matters, but only in the early years, and that young owner are less effective than older ones.

Let us now turn to the effect of a change in a variable. What would be the effect on efficiency if the secondary income from ordinary farming increased? As we can see from the formula, (6.2) the marginal effect of a marginal change in z depends on the value of the explanatory variables z. To calculate a marginal effect, it must therefore be for a specific value of z. The value could correspond to a specific firm or the mean firm. We could also calculate the effect for all firms and then take the mean. We will show how to do this and then compare the results with those achieved using the OLS model. In R, the value of the distribution function for a standardized normal distribution at the point x results from the function `pnorm(x)`, and the calculations corresponding to (6.2) are shown below:

```
# The tobit model
eTob <- tobit(E ~ zz1+z3+z6, left=-Inf, right=1, data=d)
# the standard error, needed for the use of standard normal dist.
s <- sqrt(var(residuals(eTob)))
# The mean at the effect for all firms
az <- fitted(eTob)
mean(coef(eTob)[2] * (pnorm((1-az)/s) - pnorm(-az/s)))
mean(coef(eTob)[3] * (pnorm((1-az)/s) - pnorm(-az/s)))
mean(coef(eTob)[4] * (pnorm((1-az)/s) - pnorm(-az/s)))
# the effect at the mean of az
az <- mean(fitted(eTob))
coef(eTob) * (pnorm((1-az)/s) - pnorm(-az/s))
# the effect at the min value of az
az <- min(fitted(eTob))
coef(eTob) * (pnorm((1-az)/s) - pnorm(-az/s))
# the effect at the max value of az
az <- max(fitted(eTob))
coef(eTob) * (pnorm((1-az)/s) - pnorm(-az/s))
# the OLS model
lm(E ~ zz1+z3+z6, data=d)
```

The results are collected in the Table 6.14.

If we increase z_1 with 1 000 000 and increase `zz1` by 1, then efficiency E in the OLS model will decrease by .102. In the tobit model for the firm with the lowest expected efficiency level, the minimum az, the effect on E is $-.100$, whereas for the firm with the highest efficiency level, the effect is $-.111$. If owner age increases by 10 years, efficiency increases $10 \times 0.0067 = 0.067$ for the youngest owners and

$10 \times 0.0078 = 0.0784$ for an owner of average age. Therefore, if the efficiency of a young owner is 60%, then after 10 years, it will be 66.7% ceteris paribus.

Note that the effect of the mean firm is -0.116, whereas the mean of the effect is -0.114. This is just a small difference, but it is sufficient to show that the change in efficiency is not linear in the tobit model.

6.11.4 Problems with the two-step method

The tobit model has been used in hundreds of studies of efficiency and productivity analysis but is also the focus of some recent debates.

An assumption in the model above is that z and ε are independently distributed. If that is not the case, the likelihood function might not factorize as the conditional likelihood function given z. If z and u are not independent, then we may have $\mathrm{EV}(u|z) \neq \mathrm{EV}(u)$, and many of our results above will not hold. For instance, the estimates based on the above-proposed second-stage methods might be biased and not inconsistent. An alternative is to use bootstrapping methods. Another option is to use stochastic frontier analysis (SFA), in which the relationship of dependence between efficiency and the other variables can be integrated into the model formulation by letting the mean and possibly the variance of the half-normal inefficiency term ε depend on z.

Still, theoretical niceties are of little concern to "instrumentalists", and there is considerable evidence of the success of two-stage studies in which scores are treated as descriptive measures.

Chapter 7
Stochastic Frontier Analysis SFA

7.1 Introduction

As explained in Chap. 1, there are two dominant approaches to modern benchmarking. One is the non-parametric, deterministic DEA approach discussed in some detail in the last three chapters; other is the stochastic frontier analysis (SFA), which we will cover in this and the next chapter. In the subsequent chapters, we will discuss a series of major applications of both approaches and examine the underlying benchmarking problem.

Moving from DEA to SFA, there are two main distinguishing features. One is that SFA is a *parametric approach*. By this we mean that we will make quite a few more a priori assumptions about the structure of the production possibility set and the data generation process. In fact, the SFA approach presumes that both are known a priori except for the value of a finite set of unknown parameters. While this is an obvious disadvantage of the SFA approach, it does come with a benefit. It allows us to assume a *stochastic relationship* between the inputs used and the output produced. Specifically, it allows us to assume that deviations from the frontier may reflect not only inefficiencies but also noise in the data.

In terms of methods, the DEA approach has its roots in mathematical programming, whereas the SFA approach is much more directly linked to *econometric theory*.

In this chapter, we commence our coverage of SFA by considering the simple case of a production function, that is, a multi-input single-output case. In the next chapter, we will extend our coverage to more general production settings. In this chapter, we will also focus on the estimation of SFA models and the estimation of individual efficiencies. The testing of various hypotheses is discussed in the next chapter.

P. Bogetoft and L. Otto, *Benchmarking with DEA, SFA, and R*, International Series in Operations Research & Management Science 157, DOI 10.1007/978-1-4419-7961-2_7,

7.2 Parametric approaches

Consider a production function f. Based on the technology set T, it is derived as

$$f(x) = \max\{\, y \mid (x, y) \in T \,\}$$

where x is a n dimensional input vector and y is the $m = 1$ dimensional output.

In the non-parametric DEA approach, we start with very few a priori assumptions about the production function. We may, for example, assume that f is increasing corresponding to a FDH model or that f is increasing and concave corresponding to the VRS model.

In the parametric approach, we assume a priori that the production function has a specific functional form, but that the details of this function as defined by the parameters β are unknown. That is, we assume that

$$f(x) = f(x; \beta)$$

for some unknown vector of parameters β. We may, for example, assume that the production function is a Cobb-Douglas function

$$y = \beta_0 x_1^{\beta_1} x_2^{\beta_2} \cdots x_m^{\beta_m}$$

with unknown values of $\beta_0, \beta_1, \beta_2, \ldots, \beta_m$. In the next chapter, we introduce other parametrizations more general than the Cobb-Douglas function type.

In a parametric approach, as in the non-parametric approach, we use actual observations from different firms to estimate the production function, and we use the estimated function to gauge the performance of the individual firms. More specifically, we estimate the unknown parameters β from the actual observations, (x^k, y^k), $k = 1, \ldots, K$. Let the estimated values be $\hat{\beta}$. A major difference between the parametric and the non-parametric approaches is the estimation principle. Whereas the DEA methods relied on the idea of minimal extrapolation, the parametric approaches use classical statistical principles, most notably the *maximum likelihood principle*. That is, we choose the value of $\hat{\beta}$ that makes the actual observations as likely as possible.

To implement this idea, however, we need to specify one more aspect, namely the *data generation process*, which can explain why the actual observations deviate from the production function. In the parametric approach, three main processes have been suggested. One is to consider any deviation as noise corresponding to an ordinary regression model. Another is to consider any deviation as an expression of inefficiency, much like in the DEA approach; this is called the deterministic frontier. Finally, we may assume that deviations are the results of both noise and inefficiency. This is the stochastic frontier approach.

The three approaches can be summarized as in Table 7.1, where $v \in \mathbb{R}$ is noise and $u \in \mathbb{R}_+$ is inefficiency. Consider first the *additive* specifications. We see that the noise term v can make the observed output larger or small than $f(x; \beta)$,whereas

Table 7.1 Parametric approaches to noise v and inefficiency u

Approach	Additive	Multiplicative
Regression	$y = f(x;\beta) + v$	$y = f(x;\beta)\ \exp(v)$
Deterministic	$y = f(x;\beta) - u$	$y = f(x;\beta)\ \exp(-u)$
Stochastic	$y = f(x;\beta) + v - u$	$y = f(x;\beta)\ \exp(v)\ \exp(-u)$

the inefficiency term $u \geq 0$ will always make the observed output smaller than $f(x;\beta)$. Instead of having an additive impact, we can also think of the noise and efficiency as having a *multiplicative* impact. This is often convenient, as the Farrell and Shephard efficiency measures are multiplicative by nature. Again, we see that v can both increase and lower the output, as $\exp(v) \leq 1$ when $v \leq 0$ and $\exp(v) \geq 1$ when $v \geq 0$, whereas u will always lower the output, because $\exp(-u) \leq 1$ when $u \geq 0$.

Now, once we have estimated the parametric functional form, we can also measure the output efficiencies of the individual firms. Therefore, with a given production function $f(x;\hat{\beta})$, we can evaluate the efficiency of a particular firm having used x^{o} to produce y^{o} in the additive cases by the *Farrell output efficiency*, or, which, is more common in the parametric literature, by its inverse, the *Shephard output efficiency*

$$D_o(x^{\mathrm{o}}, y^{\mathrm{o}}) = \frac{\text{Actual output}}{\text{Maximal expected output}} = \frac{f(x^{\mathrm{o}};\beta) - u^{\mathrm{o}}}{f(x^{\mathrm{o}};\beta)}. \tag{7.1}$$

If we use a multiplicative formulation instead, we retrieve similar results after a log transformation. Put differently, in the multiplicative cases illustrated in Table 7.1, we would get the Shephard output efficiency as follows:

$$D_o(x^{\mathrm{o}}, y^{\mathrm{o}}) = \frac{\text{Actual output}}{\text{Maximal expected output}} = \frac{f(x^{\mathrm{o}};\beta)\exp(-u^{\mathrm{o}})}{f(x^{\mathrm{o}};\beta)} = \exp(-u^{\mathrm{o}}). \tag{7.2}$$

In the multiplicative cases, for values D_0 close to 1, we have a particularly simple interpretation of u as

$$1 - D_o = 1 - e^{-u} \simeq 1 - (e^0 - ue^0) = 1 - (1 - u) = u,$$

where the approximation is a first-order Taylor approximation. Using the Farrell efficiency, the inverse of the Shephard efficiency, we would similarly get $F - 1 = e^u - 1 \simeq (e^0 + ue^0) - 1 = (1 + u) - 1 = u$. Therefore, the interpretation of u in the multiplicative model is that it is the relative loss in output due to the inefficiency.

Before turning to the SFA models, let us comment briefly on the simple regression and deterministic frontier models.

7.3 Ordinary regression models

When the starting point for empirical analysis is the production function, an ordinary regression technique can be used to estimate the parameters of the production function from

$$y^k = f(x^k;\beta) + v^k, \quad v^k \overset{\text{iid}}{\sim} N(0,\sigma^2), \qquad k = 1,\ldots,K.$$

If need be, we can also start from the multiplicative form and conduct the above estimation in log-linear form. The regression approach interprets all deviations from the frontier as measurement noise. The simplest way to estimate it is to assume that deviations are symmetric around zero and follow a normal distribution.

The result of such an estimation, which could be a Cobb-Douglas production function with just one input, is shown in Fig. 7.1. The estimated function in the

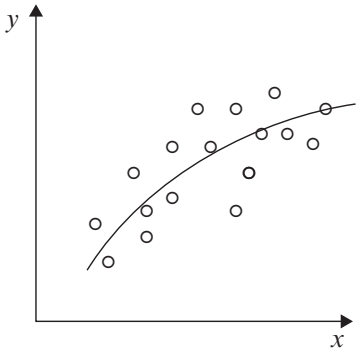


Fig. 7.1 Ordinary regression (OLS)

figure is lying more or less in the middle of all the observations, as the sum of the residuals is zero. This implies that some of the residuals are positive, where, observations lie above the estimated line, and some are negative, where observations lie below the estimated line. The estimation shown in the figure looks like one that would satisfy any statistician. The model has a clear interpretation: it is a production function, the estimated function seems to be in accordance with the observations, and the relevant parameter, the output elasticity (i.e., the relative change in output compared to the relative change in input), is apparently larger than zero. It therefore seems that we do have a good model.

There is one problem, however. Some of the observations are above the estimated production function, and this contradicts the definition of a production function that is supposed to give the maximum possible output for a given input. About half of our observations in Fig. 7.1 are above the maximum possible production.

7.4 Deterministic frontier models

If we assume instead that all deviations are the result of inefficiency, we would, as indicated above, use a model like the following:

$$y^k = f(x^k;\beta) - u^k, \quad u^k \overset{\text{iid}}{\sim} H, \qquad k = 1,\dots,K,$$

where H is some probability distribution with support only on $\mathbb{R}_+$.

We see that a deterministic parametric model assumes that there is no noise in the data like the DEA model and that the functional form—if not the specific parameters β—is given a priori. One can therefore argue that this approach comes with the drawbacks of DEA without its advantages of having a very flexible frontier specification a priori. Still, it is an interesting starting point. Historically, it can be seen as preceding the SFA approach, and moreover, a particular variant, the COLS approach, is still widely used, for example, in regulation, cf. Chap. 10.

If we assume that the functional form is linear or log-linear in β, we can estimate such a deterministic frontier model using linear programming or quadratic programming. Specifically, let us assume that

$$f(x;\beta) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m$$

we can then estimate β by solving the following problems:

$$\begin{aligned} \min_{\beta,u} \quad & \sum_{k=1}^{K} u_k \\ \text{s.t.} \quad & y^k \geq \beta_0 + \beta_1 x_1^k + \cdots + \beta_m x_m^k - u^k, \qquad k = 1,\dots,K, \\ & u^k \geq 0, \qquad k = 1,\dots,K \end{aligned} \tag{7.3}$$

or

$$\begin{aligned} \min_{\beta,u} \quad & \sum_{k=1}^{K} u_k^2 \\ \text{s.t.} \quad & y^k \geq \beta_0 + \beta_1 x_1^k + \cdots + \beta_m x_m^k - u^k, \qquad k = 1,\dots,K, \\ & u^k \geq 0, \qquad k = 1,\dots,K. \end{aligned} \tag{7.4}$$

If we assume that the inefficiency terms are exponentially distributed the first program leads to the maximum likelihood estimate β, and if the inefficiency terms are half-normal the quadratic programming problem leads to the maximum likelihood estimate of β; without any distributional assumption, the quadratic programming problem corresponds to the least square estimate of β.

A third estimation of a deterministic parametric frontier model and the one most widely used is to do a *corrected ordinary least square*, COLS . This involves two

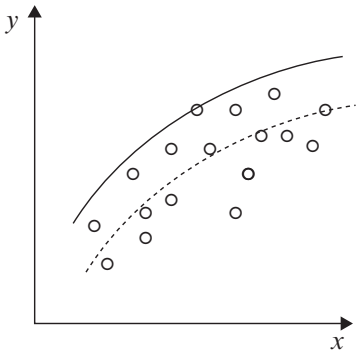


Fig. 7.2 Deterministic frontier model COLS

steps. The first is to make an ordinary least square estimate of the value of β,

$$\min_{\beta} \sum_{k=1}^{K} \left(y^k - f(x^k;\beta)\right)^2,$$

and the second is to find the smallest possible correction of the intercept β_0 to β_{00} to ensure that all observations are below the production frontier, that is, adjust β_0 upward with the maximum error term:

$$\beta_{00} = \max\{\, y^k - f(x^k;\hat{\beta}) \mid k = 1,\ldots,K \,\}.$$

An illustration of this in a log-linear case is provided in Fig. 7.2. As suggested by the theory, all of the observations are now below the estimated production function. This is also in agreement with the minimal extrapolation principle in that the corresponding technology set is the smallest set of this parametric form that contains all the data, cf. Sect. 4.6. Again, it can be shown that this leads to maximum likelihood estimates under special circumstances, namely when the inefficiency terms are gamma distributed, that is, when $u^k \sim \Gamma(\lambda,\beta)$, where $\lambda > 0$ is the shape parameter and $\beta > 0$ the scale parameter; the parameter $f = 2\lambda$ is called the degrees of freedom.

Numerical example in R

To illustrate the regressions and COLS procedures, we can use R and the data set `charnes1981` (Charnes et al, 1981). The data set is from an US federally sponsored program for providing remedial assistance to disadvantaged primary school students. The firms are 70 school sites, and data are from entire sites. The variables consists of results from three different kind of tests, a reading score, y_1, a math score, y_2, and a self–esteem score, y_3, which are considered outputs in the model, and five different variables considered to be inputs, the education level of the mother,

x_1, the highest occupation of a family member, x_2, parental visits to school, x_3, time spent with children in school-related topics, x_4, and the number of teachers at the site, x_5. There is further information in the data set, that the first 50 firms/school sites followed the program and that the last 20 are the results for sites not following the program—we will not use this important information in this example.

To make a very simple model, we choose one input variable `x1`, the mother's education, and one output variable `y1`, the reading score, and we estimate a Cobb-Douglas production function with a log-transformation

$$\log(y1) = \log(\beta_0) + \beta_1 \log(x1).$$

The following R program estimates the model by assuming implicitly that the error term is additive in the log formulation, that is, a multiplicative model in terms of Table 7.1.

```
> library(Benchmarking)
> c81 <- read.csv("charnes1981.csv")
> x = c81$x1
> y = c81$y1
> ols <- lm(log(y) ~ log(x))
> ols
Call:
lm(formula = log(y) ~ log(x))

Coefficients:
(Intercept)       log(x)
     0.9467       0.6732
> max(residuals(ols))
[1] 0.7394274
> coef(ols)[1] + max(residuals(ols))
(Intercept)
   1.686146
> plot(log(x),log(y))
> abline(coef(ols),lty="dashed")
> abline(coef(ols)[1] + max(residuals(ols)), coef(ols)[2])
> hist(exp(residuals(ols) -max(residuals(ols))), main=NULL)
```

The command `read.csv` reads the data file as a csv file, the command `lm` estimates the model as a linear model by ordinary least squares and puts the result into the object `ols`. Just entering the name of the object `ols` has the same effect as printing the object. The largest residual in this plot is 0.73, found with the command `max(residuals(ols))`. The estimate of $\beta_0 + \beta_{00}$ is 1.686.

The next three lines of code generate the graph to the left of Fig. 7.3. The first of these lines plots the individual observations, the second adds the regression line, and the last adds the COLS line.

The last line calculates the Shephard efficiencies and plots a histogram of them. The Shephard efficiency D_o was in Eq. (7.2) shown to be equal to $\exp(-u)$. The estimate of $-u$ in R is `residuals(ols) -max(residuals(ols))`, and exp of this is the efficiency. To show the distribution of the calculated efficiencies, a histogram is plotted by the command `hist` acting on the calculated efficiencies.

The results are shown in the right part of Fig. 7.3 below. From the left graph, we can

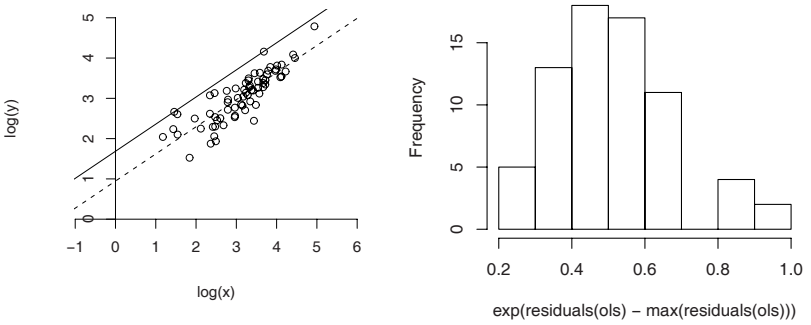


Fig. 7.3 Numerical example, COLS (*solid*) and OLS (*dashed*) lines, and histogram of COLS calculated efficiencies

see that the COLS procedure works; the adjusted estimated line is going through the upper part of the observed points.

From the right graph, we see that most firms have an efficiency value between 0.4 and 0.7. Therefore, they produce between 40% and 70% of the maximum attainable output.

7.5 Stochastic frontier models

The stochastic frontier models combine the efficiency term u with the error term v; that is, SFA models include both a stochastic error term and a term that can be characterized as inefficiency.

The base model—possibly after a log transformation—looks as follows:

$$\begin{aligned} y^k &= f(x^k;\beta) + v^k - u^k, \\ v^k &\sim N(0,\sigma_v^2),\ u^k \sim N_+(0,\sigma_u^2), \qquad k = 1,\ldots,K. \end{aligned} \tag{7.5}$$

The v term takes care of the stochastic nature of the production process and possible measurement errors of the inputs and output, and the u term is the possible inefficiency of the firm. We assume that the terms v and u are independent. If $u = 0$ the firm is 100% efficient, and, if $u > 0$, then there is some inefficiency. The N_+ denotes a half-normal distribution, i.e., a truncated normal distribution where the point of truncation is 0 and the distribution is concentrated on the half-interval $[0,\infty[$ (the support).

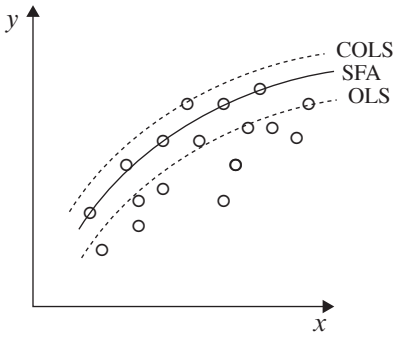


Fig. 7.4 Stochastic frontier model

An illustration of this in the usual log-linear case is provided in Fig. 7.4. If you would like to think of the model in (7.5) as a model for a Cobb-Douglas function, then y is the log of output, and x is a vector of logs of inputs including a 1 to take care of the intercept.

To estimate the SFA models, that is, to determine the values of the unknown parameters β, σ_v^2 and σ_u^2, we will use of the maximum likelihood principle. Thus, we estimate the parameter values as the values that make the observations as likely as possible. To do so, however, we must know the density of the combined error term

$$\epsilon = v - u.$$

The distribution of the error term ϵ is not a simple distribution but rather a convolution of a normal distribution, v, and a truncated normal distribution, u.

Also, although the estimated function may be of interest on its own, we are usually more interested in the resulting estimates of the individual efficiencies. That is, we would like to estimate $u^k, k = 1, \ldots, K$. An important question is therefore how to estimate them. When we estimate the model to find β, σ_v^2 and σ_u^2, we can easily calculate the total error terms

$$\epsilon^k = v^k - u^k = y^k - f(x^k; \hat{\beta}),$$

but we cannot directly obtain its components v^k and u^k.

We will now look further at the distribution of the combined error term ϵ, derive its density, and explain how to estimate the individual inefficiencies u. But first, we provide some intuition and thereafter turn to a more formal discussion of the densities and estimates.

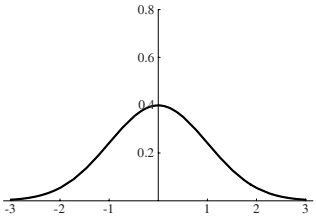


Fig. 7.5 $\sigma_v^2 = 1, \sigma_u^2 = 0, \sigma^2 = 1, \lambda = 0$

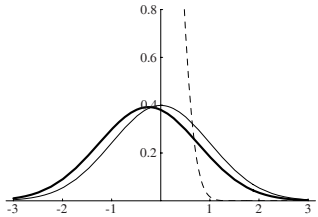


Fig. 7.5a $\sigma_v^2 = 1, \sigma_u^2 = .1, \sigma^2 = 1.1, \lambda = .3$

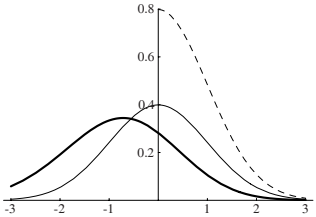


Fig. 7.5b $\sigma_v^2 = 1, \sigma_u^2 = 1, \sigma^2 = 2, \lambda = 1$

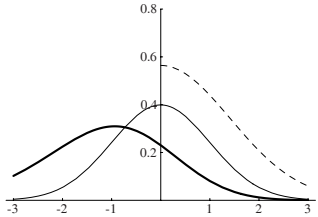
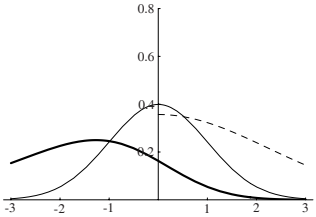


Fig. 7.5c $\sigma_v^2 = 1, \sigma_u^2 = 2, \sigma^2 = 3, \lambda = 1.4$



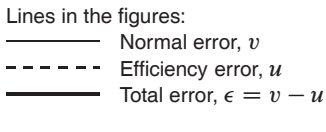


Fig. 7.5d $\sigma_v^2 = 1, \sigma_u^2 = 5, \sigma^2 = 6, \lambda = 2.2$

Fig. 7.5 The shape of the likelihood function depends on σ_v^2, σ_u^2, and λ

7.5.1 Normal and half–normal distributions

Consider the combined error term $\epsilon = v - u$, where $v \sim N(0, \sigma_v^2)$ and $u \sim N_+(0, \sigma_u^2)$, cf. Eq. (7.5) above.

If v dominates u, that is, the variance of v, σ_v^2, is much larger than the variance of u, σ_u^2, then the distribution of ϵ looks like an ordinary normal distribution; in fact, it looks like the distribution of v. If, on the other hand, u dominates v, then the distribution of ϵ looks like the distribution of u, that is, a truncated normal distribution. Of course, there are intermediate states. A series of examples are illustrated in the plots in Fig. 7.5.

In terms of parametrization, we use

$$\sigma^2 = \sigma_v^2 + \sigma_u^2 \text{ and } \lambda = \sqrt{\frac{\sigma_u^2}{\sigma_v^2}}\,.$$

When $\sigma_u^2 = 0$ and therefore $\lambda = 0$, we have the ordinary regression case with error terms like the top-left plot in Fig. 7.5.

When the variance for the inefficiency term u is positive, it follows a half-normal distribution on the positive part, which is shown as the dashed line in the plots in Fig. 7.5. The normal part is the full line, and the total error, that is, the normal error minus the efficiency term, is shown as the bold line.

If the variance of the inefficiency term is very small, that is, λ is close to 0, the density of the inefficiency term u is very narrow as the dashed line in the top right of Fig. 7.5. In such cases, it is hard to distinguish between the total error term ϵ and the normal error term v—the bold line and the normal line are almost identical in this figure.

When the variance of the inefficiency term is getting larger relative to the variance of the normal error term, that is, λ gets bigger, the density of the total error term is broader and skewed to the negative part. This is shown in the plots in Fig. 7.5. If λ is large, u is dominating, and almost all of the error term is due to differences in efficiency.

When we look at the combined error terms, we can say that a more skewed distribution indicates a greater degree to which the efficiency term dominates the normal error term. This explains how we can actually estimate the two error terms, even though they do seem to be unidentified in Eq. (7.5).

To estimate λ and σ^2, we could therefore estimate the model as an ordinary regression model, calculate the residuals, and plot their densities. We could then compare this density plot with the plots in figure 7.5 and choose the values of λ and σ^2 in these figures that look most like the plot.

Of course, this is a rather subjective method for estimating the parameters; it is also an uncertain and slow method. Moreover, once we have chosen the values for λ and σ^2, we might want to adjust the estimate of the β values. Instead, we will use a computer to make the comparisons and choose the value λ and σ^2 (and β) that makes the density curve look most like the empirical density curve of the estimated combined error terms. This can be done by the maximum likelihood estimation method. We will therefore provide a short introduction to the maximum likelihood method before we develop it for the SFA function.

7.6 Maximum likelihood estimation

Up to now, we have based our estimation of the technology set on the minimal extrapolation principle, as described in Sect. 4.3. The principle states that the technology set should be the smallest set containing all data and fulfilling certain technological assumptions such as returns to scale. We already saw in connection with Fig. 7.4 that all the data points are not below the SFA line and thus that not all data points are in the technology set derived from the parametric function. This implies that the SFA method does not fulfill the minimal extrapolation principle. This is the price we must pay to be able to handle uncertainty in the model. The consequence

is that the estimation method of the SFA model must be based on a completely different method and be motivated by other kinds of arguments.

This can be seen as a drawback of the SFA method, but it can also be seen as an advantage of the method, as it represents a way to handle uncertainty. It allows us to say that data outside the derived technology set are outside by pure chance and that these random data points should not influence what the technology set should look like.

The observations above the SFA line in Fig. 7.4 can be considered to be above by chance; either they were lucky in the production process, or there might be some measure errors of output. If we do not like observations above the estimated production line we must stay with the COLS line, where by construction there are no observations above the line. However, then we have abandoned the starting point, where we wanted to introduce uncertainty in the parametric function.

7.6.1 Justification for the method

In this section, we give an informal description of the estimation method we plan to use for SFA models.

Our stochastic frontier model has the form $\ln y = \ln f(x;\beta) + \epsilon$ where $\epsilon = v - u$, cf. Eq. (7.5). Our interest is in the unknown parameter β and u, but the parametrization in Sect. 7.5.1 shows that the parameters in the statistical model are β, σ^2, and λ—we will later show how to derive u from these parameters.

To make the writing simpler, we will only include β in what follows; that is, we make the parameters σ^2 and λ implicit in the rest of this section. And, to make it even simpler, we will treat β as if it was a scalar; readers familiar with matrix algebra can still think of β as a vector.

On the basis of our observations, we wish to decide for a value of the unknown parameter β—we want to estimate β—and we will do that by choosing a value for β such that our model is brought into agreement with our observations. The estimation of the parameter β is to choose a value that in some sense will be a good approximation of the true value of β.

Let $\varphi(y;\beta)$ be the density function for the probability distribution of y; we implicitly assume input x to make the writing simpler at this point. The density $\varphi(y;\beta)$ is normally considered to be a function of the stochastic variable y for a given parameter β (and input x). The density function $\varphi(y;\beta)$ for a given data set y can also be considered a real function on the parameter space and interpreted as the likelihood that it is the parameter β that has produced the observations y: when y is observed, β_1 is more reasonable than β_2 if $\varphi(y;\beta_1) > \varphi(y;\beta_2)$. When we consider $\varphi(y;\beta)$ as a function of the parameter β for a given y, we refer to it as the *likelihood function* $L(\beta) = \varphi(y;\beta)$. For each set of observations, we have a likelihood function, and for each parameter, we have a density function. The likelihood function can be interpreted as the likelihood or "probability" that the parameter β

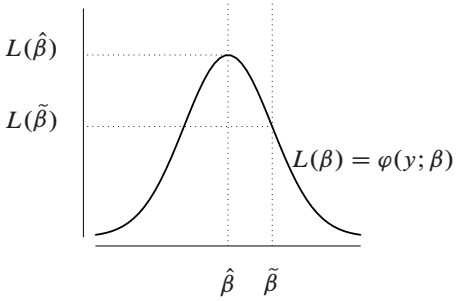


Fig. 7.6 Maximum likelihood estimate is $\hat{\beta}$ given y, the likelihood function $L(\beta) = \varphi(y;\beta)$ has its maximum at $\hat{\beta}$; the alternative $\tilde{\beta}$ has lower likelihood

has produced the observations (x, y). We often look at the *log likelihood function* $\ell(\beta) = \log\big(L(\beta)\big)$.

We choose as an estimate of β the value that maximizes the likelihood function. This is the *maximum likelihood estimation* method. We choose as a value for β the value that makes our observations the most likely observations; we choose β such that we get high agreement between our model and our observations. In mathematical terms, we choose β by maximizing $L(\beta)$. We let $\hat{\beta}$ denote the solution to this maximization problem such that for all values of β we have $L(\beta) \leq L(\hat{\beta})$, or $L(\hat{\beta}) = \max_\beta L(\beta)$.

We show this in Fig. 7.6, where the curve is the likelihood function given y and at its maximum is the maximum likelihood estimate $\hat{\beta}$. If, instead, we choose $\tilde{\beta}$ in the figure, we get a much lower likelihood.

One can show that the maximum likelihood estimates are unique and that in large samples they are nearly unbiased, consistent, i.e. the estimated parameter value will be very close to the true value of the parameter, and efficient, i.e., have variances nearly equal to the lowest possible variance (the Cramér-Rao lower bound). One can also show that maximum likelihood estimate $\hat{\beta}$ is approximately normally distributed; we return to this result in Sect. 8.5.

7.6.2 Numerical methods

Unfortunately, it is not always easy to maximize the likelihood function. For the model we have described, there is no direct closed-form solution, and we have to make use of numerical methods. We now describe the principle of such a numerical methods, and even though it can be seen as a somewhat technical issue, there are at least two reasons to be interested in this. First to understand that a SFA method

may sometimes fail to find an estimate. Second to understand how we get variances for the estimated parameters. We need the variances in the next chapter when we, among other things, look at statistical tests in SFA models.

It turns out that it is frequently easier to maximize the log likelihood function than it is to maximize the ordinary likelihood function. The solution of the maximization problem can be found as a solution to the first order conditions, often called *the likelihood equations*,

$$\frac{\partial \ell(\beta)}{\partial \beta_i} = 0, \qquad i = 1, \ldots, m.$$

For the SFA model, there is no direct solution to these equations, so we have to solve them numerically. This can be done by using a first-order Taylor expansion to the likelihood equation to obtain

$$0 = \frac{\partial \ell(\hat{\beta})}{\partial \beta} \simeq \frac{\partial \ell(\beta^0)}{\partial \beta} + (\hat{\beta} - \beta^0)\frac{\partial^2 \ell(\beta^0)}{\partial \beta^2},$$

such that

$$\hat{\beta} \simeq \beta^0 - \left(\frac{\partial^2 \ell(\beta^0)}{\partial \beta^2}\right)^{-1} \frac{\partial \ell(\beta^0)}{\partial \beta}. \tag{7.6}$$

This equation can be used in an iterative process to solve for β

$$\beta^{n+1} = \beta^n - \left(\frac{\partial^2 \ell(\beta^n)}{\partial \beta^2}\right)^{-1} \frac{\partial \ell(\beta^n)}{\partial \beta}.$$

Start with an initial guess β^0 and repeat the above formula by replacing β^n with the newly calculated value β^{n+1} from the left side.Repeat this process until β^{n+1} does not change from β^n, that is, $|\beta^{n+1} - \beta^n| < \varepsilon$ for some $\varepsilon > 0$; typically, ε could be 10^{-4}. If the value of β itself is very small, we could use the criteria $\frac{|\beta^{n+1} - \beta^n|}{\beta^n + 1} < \varepsilon$. The above method is called *Newton's method*. As an initial guess, one can use the parameters estimated by ordinary OLS or what we called COLS on page 201.

Because the likelihood function for the stochastic frontier model is non-linear in its parameters, and no closed-form solution for the parameters exists, the estimation is done by an iterative optimization algorithm, as described. Therefore, there is a chance that the optimization algorithm does not converge or returns parameters that do not correspond to the global maximum of the likelihood function.

7.7 The likelihood function

To be able to use maximum likelihood estimation in SFA, we need the likelihood function.

The density function for a single observation of one error term, v, is the normal distribution

$$\varphi_v(v) = \frac{1}{\sqrt{2\pi\sigma_v^2}}\, e^{-\frac{1}{2}\frac{v^2}{\sigma_v^2}}, \tag{7.7}$$

and the density for the inefficiency term u is the half–normal distribution, which is the normal distribution truncated at 0,

$$\varphi_u(u) = \begin{cases} \dfrac{2}{\sqrt{2\pi\sigma_u^2}}\, e^{-\frac{1}{2}\frac{u^2}{\sigma_u^2}} & \text{for } u \geq 0 \\ 0 & \text{for } u < 0 \end{cases} \tag{7.8}$$

where the extra 2-factor is such that the total mass of the half–normal distribution is still 1, that is, $\int_{-\infty}^{\infty} \varphi_u(u)\, du = 1$.

When we look at a single observation (x, y), we can not directly calculate the v and u terms. We can calculate the total error term $\epsilon = v - u$ as $\epsilon = y - f(x;\beta)$ or $\epsilon = \log y - \log f(x;\beta)$. We therefore need to find the distribution function or the density function of ϵ. The total error $\epsilon = v - u$ is the sum of v and $-u$ and therefore the distribution of ϵ is the convolution of the distribution of v and $-u$, and this is given by

$$\varphi_\epsilon(\epsilon) = \int_{-\infty}^{\infty} \varphi_u(u)\varphi_v(\epsilon + u)\, du = \int_0^{\infty} \varphi_u(u)\varphi_v(\epsilon + u)\, du. \tag{7.9}$$

Carrying out this integration is a somewhat tedious task, and the details are therefore only given in the appendix to this chapter. The result is

$$\varphi_\epsilon(\epsilon) = \frac{\sqrt{2}}{\sqrt{\pi\sigma^2}}\, \Phi\left(-\frac{\lambda\epsilon}{\sqrt{\sigma^2}}\right)\, e^{-\frac{1}{2}\frac{\epsilon^2}{\sigma^2}}, \tag{7.10}$$

where as before

$$\sigma^2 = \sigma_v^2 + \sigma_u^2, \tag{7.11}$$

$$\lambda = \sqrt{\frac{\sigma_u^2}{\sigma_v^2}}\,, \tag{7.12}$$

and Φ is the distribution function of the standard normal distribution with mean 0 and variance 1, that is, $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}t^2}\, dt$. When the parameter λ is 0, there is no effect from differences in efficiency, and if it is very large, differences are almost only due to differences in efficiency and not to other kind of uncertainty.

The log of this density is

$$\log \varphi_\epsilon(\epsilon) = -\tfrac{1}{2}\log\Big(\frac{\pi}{2}\Big) - \tfrac{1}{2}\log\sigma^2 + \log\Phi\Big(-\frac{\lambda\epsilon}{\sqrt{\sigma^2}}\Big) - \tfrac{1}{2}\frac{\epsilon^2}{\sigma^2}. \tag{7.13}$$

When we have K independent observations, K firms, the joint density is

$$\varphi(\epsilon_1, \ldots, \epsilon_K) = \prod_{k=1}^{K} \varphi_\epsilon(\epsilon_k)\,,$$

and the log of the joint density is

$$\log \varphi(\epsilon_1, \ldots, \epsilon_K) = \sum_{k=1}^{K} \log \varphi_\epsilon(\epsilon_k)$$

$$= -\tfrac{1}{2} K \log\Big(\frac{\pi}{2}\Big) - \tfrac{1}{2} K \log \sigma^2 + \sum_{k=1}^{K} \log \Phi\Big(-\frac{\lambda\, \epsilon_k}{\sqrt{\sigma^2}}\Big) - \frac{1}{2\sigma^2} \sum_{k=1}^{K} \epsilon_k^2\,.$$

We can rewrite this equation to emphasize that the error term ϵ depends on the parameter (vector) β, and then the log likelihood function looks like

$$\begin{aligned} l(\beta, \sigma^2, \lambda) &= \log \varphi_e\big(\epsilon_1(\beta), \ldots, \epsilon_K(\beta); \sigma^2, \lambda\big) \\ &= \log \varphi_e\big(y_1 - f(x_1; \beta), \ldots, y_K - f(x_K; \beta); \sigma^2, \lambda\big) \\ &= -\tfrac{1}{2} K \log\Big(\frac{\pi}{2}\Big) - \tfrac{1}{2} K \log \sigma^2 + \sum_{k=1}^{K} \log \Phi\Big(-\frac{\lambda\,\big(y^k - f(x^k; \beta)\big)}{\sqrt{\sigma^2}}\Big) \\ &\quad - \frac{1}{2\sigma^2} \sum_{k=1}^{K} \big(y^k - f(x^k; \beta)\big)^2\,. \end{aligned} \tag{7.14}$$

The function $l(\beta, \sigma^2, \lambda)$ is the *log-likelihood function*, which depends on parameters to be estimated (in this case β, σ^2, and λ) and on the data (x_1, y_1),…,(x_K, y_K).

7.8 Actual estimation

We can estimate the parameters β, σ^2, and λ of the basic SFA model in Eq. (7.5) using the maximum likelihood method. maximizing the log-likelihood function Eq. (7.14) with respect to the parameters β, σ^2, and λ. This can be done automatically by the function `sfa` from the R package Benchmarking.

Numerical example in R

We will illustrate the use of the function `sfa` on the school data that we also analyzed in Sect. 7.4.

Let us assume that data are already read into the matrices x and y and that the package Benchmarking is loaded. The following commands will then estimate a Cobb-Douglas production frontier, i.e. a log-linear production function, and display the results.

```
> msfa <- sfa(matrix(log(x)),matrix(log(y)))
> msfa
Coefficients:
(Intercept)            x
     1.2526       0.6555
```

Just as for the method `dea`, input and output to `sfa` should be in the form of matrices. We have put the output from the estimation into a variable `msfa` to be able to refer to the results at a later stage.

To compare the estimation results with the ordinary regression results from Sect. 7.4, we repeat the results here.

```
> ols <- lm(log(y) ~ log(x))
> ols
Call:
lm(formula = log(y) ~ log(x))
Coefficients:
(Intercept)       log(x)
     0.9467       0.6732
> max(residuals(ols))
[1] 0.7394274
```

The estimates slope of the two are almost identical, the OLS slope is estimated at 0.6732, and the SFA slope is estimated at 0.6555. It is not always the case that the estimated parameters from OLS and SFA are so similar, as is the case here. The differences in the estimated curves is illustrated in the plot in Fig. 7.7.

The commands in Fig. 7.7 plot the observations as well as the production functions estimated by OLS (dashed line), OLS plus max residual (dotted line), and SFA (continues line). The production frontier estimated by SFA (the solid line) is clearly

```
> plot(log(x), log(y))

> abline(coef(msfa))
> abline(coef(ols),
+ lty = "dashed")
> abline(coef(ols)[1] +
+ max(residuals(ols)),
+ coef(ols)[2],
+ lty="dotted")
```

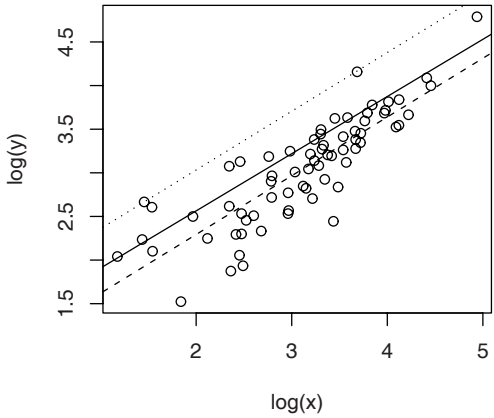


Fig. 7.7 Plot of SFA line (solid); also OLS (*dashed*) and COLS (*dotted*) lines

above the function estimated by OLS (the dashed line) and also below the OLS line plus the maximum residual (the dotted line), but the difference in the slope of the

SFA line (the solid line) and the OLS line (the dashed line) do not clearly correspond to the similarity in the slopes we mentioned above to a great degree. There are still observations above the production frontier, but apparently these deviations are due to random errors (v).

The estimate for λ can be found by either `msfa$lambda` or the function `lambda.sfa(msfa)`.

```
> lambda.sfa(msfa)
    lambda
1.191951
```

The percentage of total error variance due to inefficiency can be found from $\frac{\sigma_u^2}{\sigma_u^2+\sigma_v^2} = \frac{\sigma_u^2/\sigma_v^2}{\sigma_u^2/\sigma_v^2+\sigma_v^2/\sigma_v^2} = \frac{\lambda^2}{\lambda^2+1}$. For $\hat{\lambda} = 1.192$, we get $\frac{\lambda^2}{\lambda^2+1} = 0.5869$, showing that 59% of the total variation is due to inefficiency and that the remaining 41% is random variation.

7.9 Efficiency variance

The estimation procedure generates estimates of λ and σ^2, but our interest is often rather in σ_u^2 and σ_v^2. They can be found by solving for σ_u^2 and σ_v^2 in the Eq. (7.11), $\sigma^2 = \sigma_v^2 + \sigma_u^2$, and Eq. (7.12), $\lambda = \sqrt{\frac{\sigma_u^2}{\sigma_v^2}}$. This can be done in a rather straightforward manner:

$$\lambda = \sqrt{\frac{\sigma_u^2}{\sigma_v^2}} \quad \Rightarrow \quad \lambda^2 = \frac{\sigma_u^2}{\sigma_v^2} \quad \Rightarrow \quad \sigma_u^2 = \lambda^2\sigma_v^2.$$

We now can find

$$\sigma^2 = \sigma_v^2 + \sigma_u^2 = \sigma_v^2 + \lambda^2\sigma_v^2 = \sigma_v^2(1+\lambda^2) \quad \Rightarrow \quad \sigma_v^2 = \frac{1}{1+\lambda^2}\,\sigma^2$$

and therefore

$$\sigma_u^2 = \lambda^2\sigma_v^2 = \frac{\lambda^2}{1+\lambda^2}\,\sigma^2.$$

Applying these equations to our estimates $\hat{\lambda} = 1.1920$ and $\hat{\sigma^2} = 0.1663$, we get $\sigma_u^2 = 0.099$ and $\sigma_v^2 = 0.070$. As $\hat{\lambda}$ is greater than 1, the variance for efficiency is larger than the variance for random errors. The two variances can be found directly in R:

```
> sigma2u.sfa(msfa)
  sigma2u
0.09877588
> sigma2v.sfa(msfa)
  sigma2v
0.06952391
```

The two variances are also part of the output from the command `summary(msfa)` where the variable `msfa` is the name of the sfa-object from the `sfa`-function. Note that the variance for the inefficiency σ_u^2 is around 42% larger than the variance for the random error σ_v^2; this corresponds to $\lambda^2 = 1.1920^2 = 1.420$ and then $1.420 - 1 = 42\%$.

Practical application: Milk producers

We estimate a simple model for milk production, where output is kg milk and inputs are veterinary expenses, energy and the number of cows. The commands below read the data, which are in csv format, show a part of the data set, and estimate a Cobb-Douglas production frontier using OLS (`lm`) and SFA.

```
> library(Benchmarking)
> milkProd <- read.csv( "milkProd.csv" )
> milkProd[ c( 1:3, 107:108 ), ]
    farmNo   milk energy    vet cows
1        1 862533 117894  21186  121
2        2 605764  72049  43910   80
3        3 865658 158466  54583   95
107    107 983645 190440  73142  116
108    108 738916 156109 115209   92
> x <- with(milkProd, cbind(vet, energy, cows))
> y <- matrix(milkProd$milk)
> milkSfa <- sfa(log(x), log(y))
> summary(milkSfa)
            Parameters   Std.err   t-value  Pr(>|t|)
(Intercept)    7.52014   0.32197    23.357     0.000
xvet           0.06281   0.02496     2.517     0.013
xenergy        0.12156   0.03676     3.307     0.001
xcows          0.87879   0.06640    13.235     0.000
lambda         3.59708   0.89964     3.998     0.000
sigma2         0.045685
sigma2v =  0.003277507 ;  sigma2u =  0.04240755
log likelihood =  67.82555
>
> # Percentage of inefficiency variation to total variation
> lambda <- lambda.sfa(milkSfa)
> 100*lambda^2/(1+lambda^2)
92.82587
> # variance for inefficiency
> sigma2u.sfa(milkSfa)
   sigma2u
0.04240755
> # variance for random errors
> sigma2v.sfa(milkSfa)
    sigma2v
0.003277507
```

In this model, the estimated λ parameter is 3.6, which means that the total error variance is mainly due to inefficiency, whereas random errors are less important. The percentage of total variation due to variation in efficiency is 93%.

The estimated variance for the variation in efficiency is $\sigma_u^2 = 0.0424$ is considerably larger than variation due to random errors $\sigma_v^2 = 0.0032$.

7.9.1 Comparing OLS and SFA

The following lines of code compare the OLS and the SFA estimates of the parameters in the model.

```
> library(Benchmarking)
> milkProd <- read.csv( "milkProd.csv" )
> x <- with(milkProd, cbind(vet, energy, cows))
> y <- matrix(milkProd$milk)
>
> # sfa efficiency
> milkSfa <- sfa(log(x), log(y))
> ols <- lm(log(y) ~ log(x))
> cbind(ols=coef(ols), sfa=coef(milkSfa))

                     ols        sfa
(Intercept)  7.10341187 7.52014420
log(x)vet    0.09551563 0.06281416
log(x)energy 0.12132193 0.12156101
log(x)cows   0.85907831 0.87878814
> # ols variance
> sum(residuals(ols)^2)/ols$df.residual
[1] 0.02046301
> max(residuals(ols))
[1] 0.2817728
```

The R command `cbind` writes the parameters for the two models, OLS and SFA, and here one can see that the estimates are different. However, except for the intercept, the differences are rather small. The corrected intercept for COLS is the intercept plus the maximum of the residuals, that is, $7.10 + 0.28 = 7.38$, and this intercept is much closer to the SFA intercept. This is an often-found result, as the OLS estimates differ only a little from the SFA estimates. On the other hand, the OLS variance 0.02046 is much smaller than the total SFA variance of 0.04568 (`sigma2`); this is no surprise, as the OLS estimates are found such that the variance is the lowest attainable variance, and this leads to its name as the least squares or OLS approach.

7.10 Firm-specific efficiency

So far, we have focused on the estimation of the functional form and whether the deviations from the production function can be decomposed into noise and inefficiency. We have not, however, analyzed the efficiency of individual firms, which, after all, is the major concern in benchmarking studies. We therefore turn to firm-specific efficiency in SFA analysis.

The calculations in this section might look tedious, but they represent the foundation for calculating the specific efficiency for each firm. We will show also how it can be done in R; in fact, the R package contains specialized functions that can calculate the specific efficiencies in one function call after the SFA estimation has been done, just as for the DEA efficiencies.

The efficiency of the specific firm in both the additive and multiplicative model depends on u. In the multiplicative model, the efficiency depends only on u, Eq. (7.2), and, in the additive model, the efficiency also depends on the maximal expected output, that is, the output determined from the estimated function, Eq. (7.1). The firm-specific efficiency is therefore given by

$$TE^k_{\text{add}}(x^k, y^k) = \frac{f(x^k, \hat{\beta}) - \hat{u}^k}{f(x^k, \hat{\beta})} = 1 - \frac{\hat{u}^k}{f(x^k, \hat{\beta})}, \tag{7.1'}$$

$$TE^k = TE^k_{\text{mult}}(x^k, y^k) = \exp(-\hat{u}^k)\,. \tag{7.2'}$$

where technical efficiency TE without a subscript refers to the multiplicative model. Whatever model we use, we need an estimate $\hat{u}^k$ to be able to calculate the specific efficiency. Unfortunately, it is not simple to get an estimate $\hat{u}^k$ of u^k. After estimating the parameters, we can easily estimate the total error as

$$\hat{\epsilon}_k = \ln y_k - \ln f(x_k; \hat{\beta})\,, \qquad k = 1, \ldots, K.$$

The total error is given by $\epsilon_k = v_k - u_k$ from Eq. (7.5), but even though we know ϵ_k, this is one equation and two unknowns, v_k and u_k.

Still, the estimate of ϵ_k does carry some information on u_k. If $\epsilon_k > 0$, then chances are that u_k is not very large, as $\text{EV}(v_k) = 0$ and $u_k \geq 0$, suggesting that firm k is relatively efficient. If, on the other hand, $\epsilon_k < 0$, then u_k will tend to be large, suggesting that firm k is relatively inefficient.

We will therefore look at the conditional distribution of u_k given ϵ_k and use the conditional expectation $\text{EV}(u_k|\epsilon_k)$ as an estimator of u^k. The simultaneous density of v and u—we drop the subscript k for a moment—is the product of the individual densities, as they are independent $\varphi_{u,v}(u, v) = \varphi_u(u)\,\varphi_v(v)$. Substituting $\epsilon + u$ for v, we get $\varphi_{u,\epsilon}(u, \epsilon) = \varphi_v(\epsilon + u)\,\varphi_u(u)$. Therefore, using Bayes' theorem, the conditional density of u given ϵ is

$$\varphi(u|\epsilon) = \frac{\varphi_v(\epsilon + u)\,\varphi_u(u)}{\varphi_\epsilon(\epsilon)}$$

where φ_v, φ_u, and $\varphi_\epsilon(\epsilon)$ are defined as in Eqs. (7.7), (7.8) and (7.10) respectively. Unfortunately, the actual calculations are rather tedious, and as we simply aim to find the conditional expectation, we will jump right to this result:

$$\mathrm{EV}(u|\epsilon) = \mu_* + \sigma_* \frac{\phi(\mu_*/\sigma_*)}{\Phi(\mu_*/\sigma_*)}, \tag{7.15}$$

where

$$\mu_* = -\epsilon \frac{\sigma_u^2}{\sigma^2} = -\epsilon \frac{\lambda^2}{1+\lambda^2} = -\epsilon\gamma,$$
$$\sigma_* = \sqrt{\frac{\sigma_u^2\sigma_v^2}{\sigma^2}} = \frac{\lambda}{(1+\lambda^2)}\sigma = \sqrt{\gamma(1-\gamma)\sigma^2}\,,$$

and $\phi(.)$ is the density function, and $\Phi(.)$ the distribution function of a standard normal distribution. When we substitute the estimated values for ϵ, σ^2, and λ then we have an estimate of u, call it $\hat{u}$, conditioned on the estimate of ϵ.

For the multiplicative model, we now get an estimate of TE as $\widehat{TE} = e^{-\mathrm{EV}(u|\hat{\epsilon})} = e^{-\hat{u}}$. The following commands extract the residuals ϵ, σ^2, and λ.

```
> e <- residuals(milkSfa)
> s2 <- sigma2.sfa(milkSfa)
> lambda <- lambda.sfa(milkSfa)
```

Now, we can calculate the auxiliary variables μ_* and σ_* and the specific technical efficiency estimates of each firm.

```
> mustar <- -e*lambda^2/(1+lambda^2)
> sstar <- lambda/(1+lambda^2)*sqrt(s2)
> teJ <- exp(-mustar
       -sstar*( dnorm(mustar/sstar)/pnorm(mustar/sstar) ))
```

We can also note that

$$\frac{\mu_*}{\sigma_*} = -\epsilon \frac{\sigma_u^2}{\sigma_2^2} \frac{\sigma}{\sigma_u\sigma_v} = -\epsilon \frac{\sigma_u}{\sigma_v}\frac{1}{\sigma} = -\epsilon\frac{\lambda}{\sigma} \quad \text{where} \quad \lambda = \frac{\sigma_u}{\sigma_v}$$

such that

$$\mathrm{EV}(u|\epsilon) = \sigma_* \left(\frac{\phi(\epsilon\lambda/\sigma)}{1-\Phi(\epsilon\lambda/\sigma)} - \epsilon\frac{\lambda}{\sigma} \right) \tag{7.16}$$

where we have used that $\phi(-x) = \phi(x)$ and $\Phi(x) = 1 - \Phi(-x)$. The above equation can be simplified to

$$\mathrm{EV}(u|\epsilon) = \sigma_* \left(\frac{\phi(\epsilon_*)}{1-\Phi(\epsilon_*)} - \epsilon_* \right) \quad \text{where } \epsilon_* = \epsilon\frac{\lambda}{\sigma}\,. \tag{7.17}$$

The following commands calculate the technical efficiency estimates using the above formula and show that these estimates are equal to the estimates calculated in Eq. (7.15).

```
> estar <- e * lambda / sqrt(s2)
> euJ <- sstar * (dnorm(estar)/(1 - pnorm(estar)) - estar)
> teJJ <- exp(-euJ)
> all.equal(teJ, teJJ)
[1] TRUE
```

Another estimator is the mode of the conditional distribution, which also can be interpreted as a maximum likelihood estimator:

$$M(u|\epsilon) = \begin{cases} \mu_* & \text{for } \epsilon \leq 0, \\ 0 & \text{for } \epsilon > 0. \end{cases} \tag{7.18}$$

Because μ_* always has the opposite sign of ϵ, we can change the above equation to

$$M(u|\epsilon) = \begin{cases} \mu_* & \text{for } \mu_* > 0, \\ 0 & \text{for } \mu_* \leq 0. \end{cases} \tag{7.19}$$

so that we have

$$M(u_k|\epsilon) = \max(0, \mu_{*i}). \tag{7.20}$$

```
> teMode <- exp(pmin(0,-mustar))
```

As $\mathrm{EV}(TE) = \mathrm{EV}(e^{-u})$ is generally not equal to $e^{-\mathrm{EV}(u)}$ yet another estimator has been proposed

$$TE = \mathrm{EV}(e^{-u}|\epsilon) = \frac{\Phi\left(\mu_*/\sigma_* - \sigma_*\right)}{\Phi(\mu_*/\sigma_*)} e^{\left(\frac{1}{2}\sigma_*^2 - \mu_*\right)}. \tag{7.21}$$

This estimator is optimal in the sense of minimizing the mean square error. This is the one that is most often used and the one that we will also use.

```
> teBC <- pnorm(mustar/sstar -sstar)/pnorm(mustar/sstar) *
+    exp(sstar^2/2 -mustar)
```

The actual values of the efficiency estimates differ somewhat between the three methods, but the estimates based on the three different methods are highly correlated.

```
> cor(cbind(teBC=c(teBC), teMode=c(teMode), teJ=c(teJ)))
             teBC    teMode       teJ
teBC   1.0000000 0.9955548 0.9999965
teMode 0.9955548 1.0000000 0.9957821
teJ    0.9999965 0.9957821 1.0000000
```

The average efficiency for the industry is the average over the individual production firms $\overline{TE} = \frac{1}{N}\sum_{i=1}^{N} \widehat{TE}_k$.

```
> mean(teBC)
[1] 0.858807
```

However, it is worth consideration whether one should weight the average by the output, that is, use

$$\sum_{i=1}^{N} \frac{y_k}{\sum_{j=1}^{N} y_j} \widehat{TE}_k = \frac{\sum_{i=1}^{N} \widehat{TE}_k \; y_k}{\sum_{i=1}^{N} y_k}. \tag{7.22}$$

```
> sum(milkProd$milk*teBC/sum(milkProd$milk))
[1] 0.8709019
```

The calculation of the efficiency estimates based on the above formulas can be simplified by applying the methods `te.sfa` or `teBC.sfa` for teBC, `teMode.sfa`, and `teJ.sfa` on objects returned by the R command `sfa`. The next couple of lines show that the results are the same.

```
> all.equal(matrix(teBC),te.sfa(milkSfa),check.attributes=FALSE)
[1] TRUE
> all.equal(matrix(teMode),teMode.sfa(milkSfa),check.attributes=F)
[1] TRUE
> all.equal(matrix(teJ),teJ.sfa(milkSfa),check.attributes=FALSE)
[1] TRUE
```

The following commands plot two graphs that visualize the efficiency estimates calculated with the formula from (7.21). They are shown in Fig. 7.8.

```
hist(te, xlim = c(0.5, 1), main = "", xlab = "Efficiency",
    col = "gray", cex.lab = 1.25, freq=F, breaks=10)
lines(density(te,from=0,to=1),lwd=2)

plot(sort(teBC), ylim = c(0.5, 1), ylab = "Efficiency", pch = 20,
     cex.lab = 1.25)
```

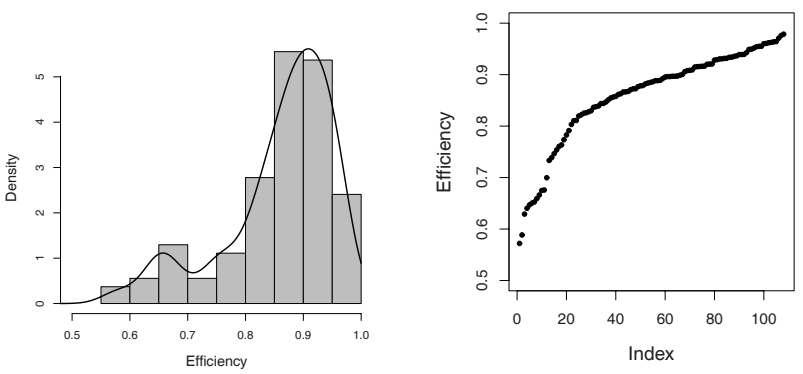


Fig. 7.8 Efficiencies: Histogram, density, and order

Finally, we look at the relationship between the production of milk and efficiency. The following commands are used to plot the graph shown in Fig. 7.9.

```
plot(milkProd$milk, teBC, xlab = "Kg milk produced",
     ylab = "efficiency")
lines(lowess(milkProd$milk, teBC), lwd = 1.5)
```

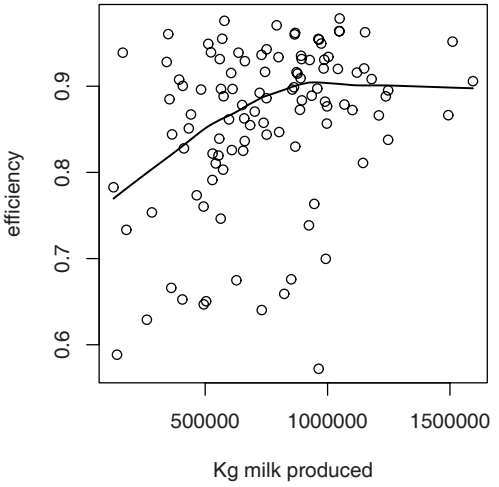


Fig. 7.9 Relationship between milk production and efficiency

This graph shows that efficiency increases with the production of milk for the firms that are smaller than the average size but that efficiency is independent of milk production for the firms that are larger than the average size.

7.10.1 Firm-specific efficiency in the additive model

For the additive model, the efficiency was shown to be

$$TE^k_{\text{add}}(x^k, y^k) = 1 - \frac{\hat{u}^k}{f(x^k, \hat{\beta})}, \qquad k = 1, \ldots, K, \tag{7.1"}$$

and estimates of u were found in (7.16) and (7.18). We thus have all the ingredients to calculate firm-specific efficiency in the additive model. The following lines of code show how this is done in R. We continue to use the data set for milk production we introduced in Sect. 7.9.

The first step is to estimate the additive model. This is almost like the estimate of the multiplicative model; we just omit taking the log of the variables. In the second step, we estimate efficiencies by the function `te.add.sfa`.

```
> library(Benchmarking)
> milkProd <- read.csv( "milkProd.csv" )
```

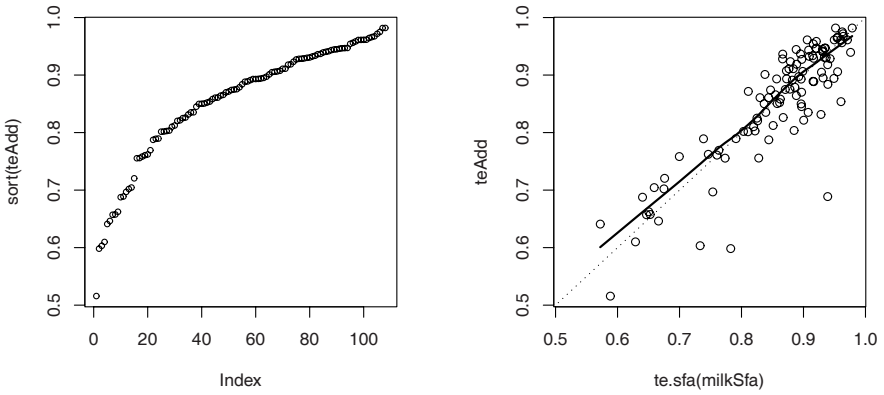


Fig. 7.10 Efficiencies in additive model sorted and compared to efficiencies in the multiplicative model

```
> x <- with(milkProd, cbind(vet, energy, cows))
> y <- matrix(milkProd$milk)
> milkAdd <- sfa(x, y)
> milkAdd
Coefficients:
(Intercept)         xvet     xenergy        xcows
      51156       1.1176      0.53837       6774.9
> lambda <- lambda.sfa(milkAdd)
> 100*lambda^2/(1+lambda^2)
  lambda
89.04813
> teAdd <- te.add.sfa(milkAdd)
> plot(sort(teAdd))

> # Compare to the multiplicative model:
> milkSfa <- sfa(log(x), log(y))
> ran <- range(te.sfa(milkSfa), teAdd)
> plot(te.sfa(milkSfa),teAdd , xlim=ran, ylim=ran)
> abline(0,1, lty="dotted")
> lines(lowess(te.sfa(milkSfa),teAdd),lty="solid", lwd=2)
> # Efficiency and size of output
> plot(y, teAdd, xlab="Kg milk produced",
+                ylab="Additive efficiency")
> lines(lowess(y, teAdd))
```

The sorted efficiencies `teAdd` are plotted to the left in figure 7.10. The overall impression is just like the one for the multiplicative model back in Fig. 7.8.

One cannot compare the estimates in the multiplicative model and the additive model; the first parameters correspond to output elasticities, without dimensions, and the other parameters correspond to marginal products, where dimensions depend on the inputs. However, one can compare the estimated specific efficiencies.

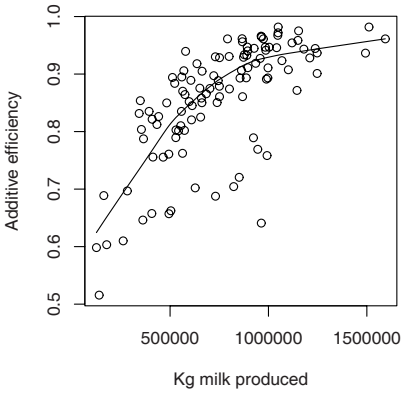


Fig. 7.11 Efficiency in additive model depends on size of production

This is done in figure 7.11. Here, one can see that there is a tendency that for low efficiencies, additive model efficiencies are larger than the multiplicative efficiencies and that for high efficiencies, the multiplicative are somewhat larger.

In Fig. 7.11 the efficiency in the additive model is compared to the size of the output, the production of milk. Here, it is clear that the dependence of size is more pronounced than for the multiplicative model in figure 7.9. However, this is not just a characteristic of the data; it actually is a characteristic of the model itself.

Efficiency in the additive model $y = \beta_0 + \beta_1 x_1 + v - u$, $v \sim N(0, \sigma_v^2)$ and $u \sim N_+(0, \sigma_u^2)$ is given by $TE_{\text{add}} = 1 - \frac{\hat{u}}{\hat{y}}$. The assumption for the stochastic model is that the distribution of u is identical for all firms. Therefore, for firms with a large maximal expected output, large $\hat{y}$, the term $\frac{\hat{u}}{\hat{y}}$ will be small, and therefore the efficiency will be close to 1. By construction, larger firms will therefore have a tendency to have a higher efficiency. If the efficiency is independent of size, then σ_u^2 should increase with size, that is, u should be heteroskedastic.

7.11 Comparing DEA, SFA, and COLS efficiencies

We can now compare the efficiencies estimated in the SFA model with the corresponding efficiencies in a DEA model. As we mentioned in Sect. 7.2, we should compare the inverse output efficiency to the SFA efficiency. The following code lines estimate efficiency under DEA and under SFA and make various graphical comparisons.

```
> library(Benchmarking)
> milkProd <- read.csv( "milkProd.csv" )
> x <- with(milkProd, cbind(vet, energy, cows))
```

```
> y <- matrix(milkProd$milk)
> # sfa efficiency

> milkSfa <- sfa(log(x), log(y))
> teSfa <- te.sfa(milkSfa)
> # dea efficiency
> milkDea <- dea(x,y,ORIENTATION="out")
> teDea <- 1/eff(milkDea)
> # COLS efficiency
> ols <- lm(log(y)~log(x))
> cols <- -residuals(ols) + max(residuals(ols))
> teCols <- exp(-cols)
> # correlation between dea, sfa, and cols efficiencies
> cor(cbind(teDea, teSfa, teCols))
                F        te    teCols
F       1.0000000 0.7790146 0.7981702
te      0.7790146 1.0000000 0.9699392
teCols 0.7981702 0.9699392 1.0000000
> plot(teDea, teSfa, xlim=c(.5,1), ylim=c(.5,1))
> boxplot(cbind(teDea, teSfa, teCols))
```

The correlation between the dea and sfa efficiencies as 0.78; that is, the two kind of efficiencies are highly correlated, but they are not perfectly correlated. This can also be seen in the left panel of figure 7.12, where there is a clear positive slope in the connection in the points.

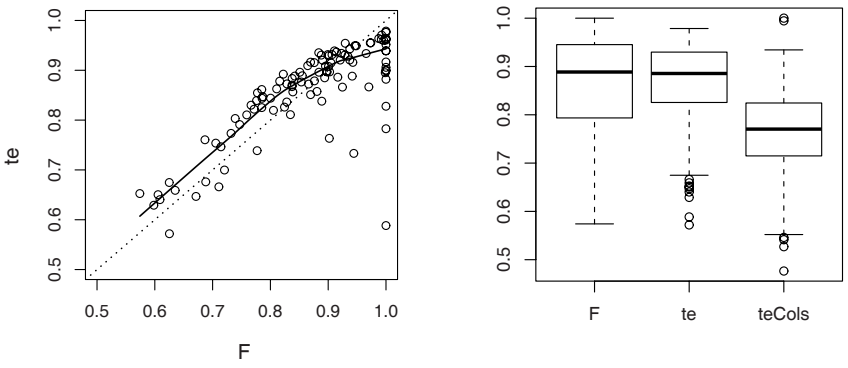


Fig. 7.12 Comparing DEA (F) and SFA (te) efficiencies, and COLS (teCols)

However, it is also clear that there are several firms with a DEA efficiency of 1 that have much lower SFA efficiency. There is even a firm with an DEA efficiency of 1.0 and a SFA efficiency of 0.6. In the graph, the diagonal is drawn (dotted line); also, a smooth line through the points is drawn (solid line). From this, one can see a tendency for the DEA efficiency to be higher than the SFA efficiency on the lower

end of the efficiency scale and for the opposite to occur on the very high end of the efficiency scale; from several data sets, this seems to be a general property and not just one of this example; cf. figure 7.13, where plots such as the left plot in figure 7.12 are given for many data sets.

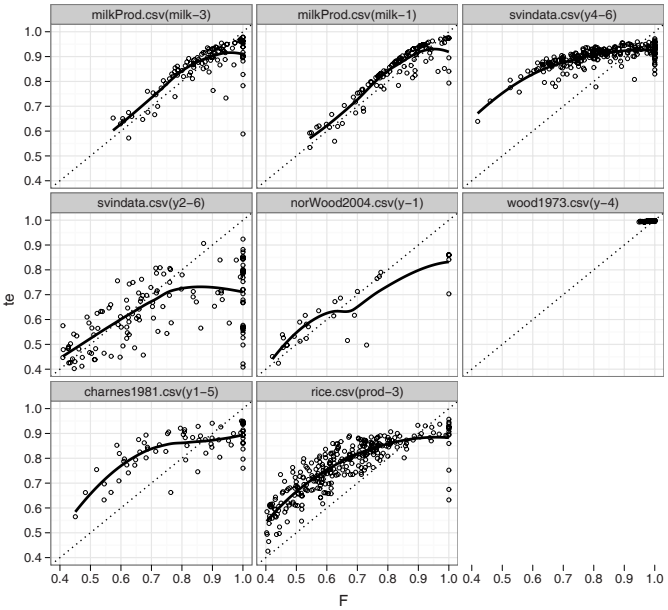


Fig. 7.13 Compare DEA efficiency (F) and SFA efficiency (te) for many data sets

As long as it is only the firms whose efficiency we calculate that determine the technology set, it is clear that at least one firm has a DEA efficiency of 1 and is fully efficient, and often we have several firms with an efficiency of 1 dependent on the total number of inputs and outputs in the model. This is not the case for SFA efficiency where an efficiency of 1 only happens when $u = 0$, and as the distribution of u is continuous, the probability of this is 0—there is no atom at 0 in the distribution, and, as a result, we do not see a gathering of u's at 0.

The firm with a DEA efficiency of 1 and a SFA efficiency of .59 is a firm with a very low use of veterinary services, an input with a small weight compared to the other inputs. Therefore, from the DEA point of view the firm is very effective in its use of an input and is therefore considered to have a high efficiency. On the other hand, this firm have a very low output milk per input cow and per input energy unit, and as the cows and energy are the two important inputs, the SFA judge the firm to be very inefficient. If veterinary services is removed from the inputs, then the DEA efficiency of that firm changes to .55 and the SFA efficiency changes to .54, leading the two to be almost the same; other differences remain more or less unchanged.

This example shows that one should be very careful in selecting inputs and outputs especially in DEA analysis—a result we also discussed in Sect. 4.6.

The right panel in Fig. 7.12 is a boxplot of the three efficiencies, F for the inverse DEA output efficiency, te for the SFA efficiency, and teCols for the COLS efficiency. Here, it is clear that even though the mean and median are pretty much the same for DEA and SFA, the spread in the DEA efficiencies is much larger than the spread in the SFA efficiencies. It is also clear that the median is lower for COLS efficiency and that there are only a few with very high efficiency. The relation between SFA and COLS efficiency is also shown in Fig. 7.14, where it is clear that for almost all firms the COLS efficiency is lower than SFA efficiency except for a few with very high COLS efficiency. This is not a surprise, as the COLS efficiency is constructed such that at least one firm has an efficiency of 1, which corresponds to the firm with the largest OLS error.

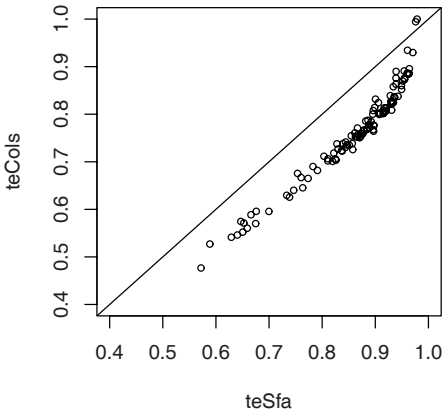


Fig. 7.14 Comparing SFA and COLS efficiencies

Let us take another look at the relation between the size of output, the amount of milk production and efficiency. In Fig. 7.9, we plotted the relation between kg milk produced and SFA efficiency. In Fig. 7.15, we have repeated this plot but now also include the DEA efficiency as both the input efficency (the dashed line) and the inverse output efficiency (the dotted line). For the middle group, the relationship is independent of how efficiency is measured. However, for the two ends of the production range, we see a different pattern, where the DEA efficiency is larger than the SFA efficiency and that DEA efficiency has a tendency to be higher at the ends of the output range. This pattern should not come as a surprise, as at the ends of the output range, there are typically fewer firms, there are few very small firms and there are few very large firms. Therefore, when the technology is a VRS technology, the firms at the ends are typically compared to very few firms, making it easier to attain a higher efficiency.

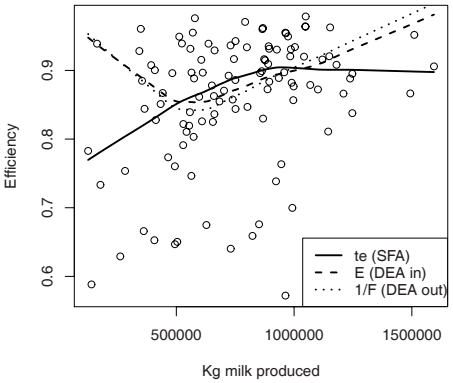


Fig. 7.15 Relationship between kg milk production and various efficiencies

Again, this seems to be a general property and not just a property for this example, as can be seen in Fig. 7.16 When we use the CRS technology for DEA efficiency, the pattern is quite different, as can be seen by comparing figure 7.16 with VRS technology with figure 7.17 with CRS technology. For the CRS plots, there is no clear pattern; note that the SFA model does not assume CRS but that the DEA and SFA efficiencies show the same pattern anyway. The conclusion is thus that one should be careful and not draw any conclusion that does not depend solely on the data but mostly on the method used.

We close this section by reminding the reader that the methodology differs between DEA and SFA, even though the two approaches yield efficiency measures. If the input is changed for an inefficient firm then that will not change the efficiency of other firms in DEA, but it might change the efficiency of other firms in SFA because it might influence what is considered random error and what is considered a difference in efficiency. Also, if more firms are added to the data set, then efficiency in DEA will only change if the new firms change the frontier; in SFA, efficiency will surely change again because the distinction between random errors and inefficiency will change. When more goods, inputs and/or outputs, are added, then an increasing number of firms will get DEA efficiency of 1; this is not necessarily the case for SFA.

7.12 Summary

In this chapter, we introduced the basics of the parametric approach to benchmarking and briefly discussed two ways in which inefficiency can enter: the additive way and the multiplicative way, which is more in accordance with the DEA approach. We showed that the ordinary regression models (OLS) do not take inefficiency into

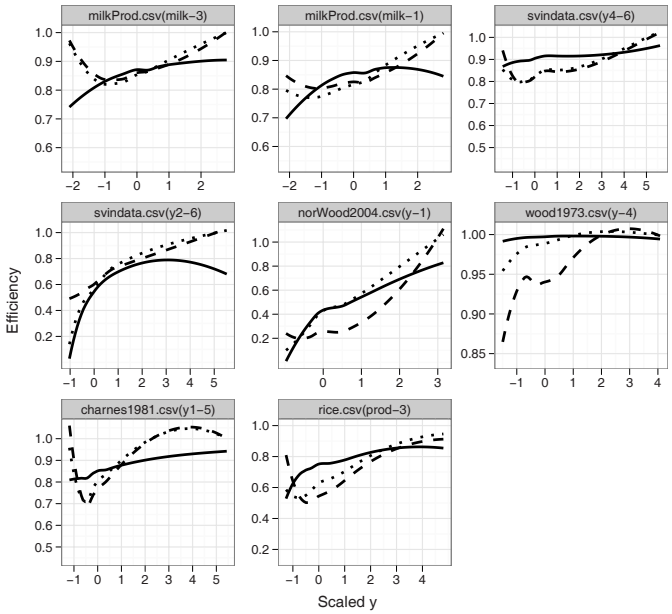


Fig. 7.16 Relation between output (scaled) and DEA efficiency (E and 1/F, *dashed* and *dotted line*) and SFA efficiency (te, *solid line*) using many data sets

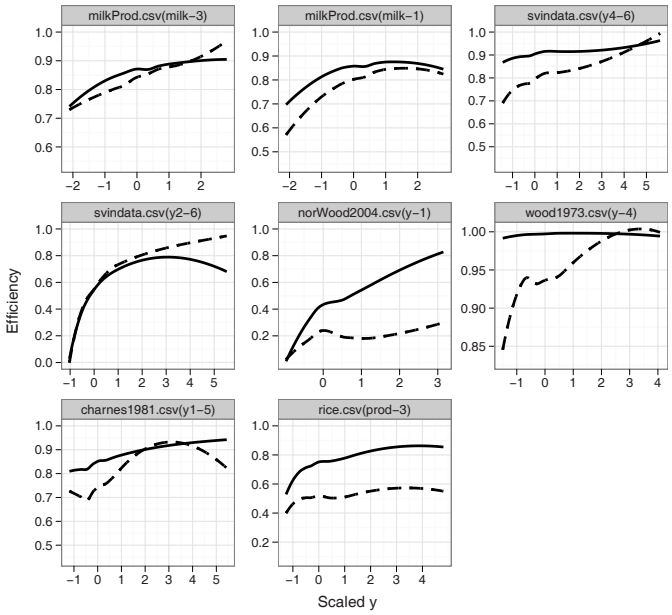


Fig. 7.17 Relation between output (scaled) and DEA efficiency (CRS technology, *dashed line*) and SFA efficiency (*solid line*) using many data sets

account at all but consider all deviations from the parametric function to be of a purely random nature. At the other extreme, the deterministic frontier models assume that all deviations are considered to be differences in efficiency; this leads to corrected ordinary least squares (COLS).

When deviations from the parametric form are split into both random errors and differences in efficiencies, we have the stochastic frontier analysis model (SFA). We introduced the distribution of efficiencies as half-normal distributions and showed what this implied for the whole stochastic model in a series of plots.

The nature of the SFA estimation problem is of a different type than the DEA method because of both the parametric functional form and the stochastic in this functional form. For estimation purposes of SFA, we introduced in a fairly informative way the maximal likelihood estimation method. We not only estimated parameters in the functional form but also showed how firm-specific efficiencies can be calculated.

At the end of the chapter, we compared SFA efficiencies to DEA efficiencies and found that almost no SFA efficiency was 1, where typically several DEA efficiencies are 1. We also noted that the methodologies in DEA and in SFA are very different and that is an important reason for the different results.

7.13 Bibliographic notes

The linear and quadratic programming approach to estimate a deterministic frontier is due to Aigner and Chu (1968). The maximum likelihood interpretation under exponential and half-normal distributions was first demonstrated by Smith (1976), Afriat (1972) proposed a gamma-distributed inefficiency distribution and Richmond (1974) noted the equivalence to a COLS approach.

A non-technical overview of maximum likelihood estimation can be found in Silvey (1970), a more mathematical discussion is found in Lehmann (1983) and Rao (1973), and a broader overview can be found in Cox and Hinkley (1974).

The use of truncated normal distribution has a long history. At the least, it is described by Anders Hald in his Danish textbook *Statistiske metoder* from 1948, translated to English in 1952 as *Statistical Theory with Engineering Applications*, with a reference to R.A. Fisher from 1931. Its introduction to econometrics was in the late seventies.

The two books, Coelli et al (1998b) and Kumbhakar and Lovel (2000), contain further references and some historical remarks on their use in econometrics. The first book is a broader description that also includes DEA. The second book is more detailed and, specializing in stochastic frontiers, has much more theory and discussion, but no empirical applications.

The book Eldén et al (2004) is a good modern introduction to numerical computations; it does not have anything on optimization, but it does contains a lot that is relevant in connection to numerical optimization as described in Frandsen et al (2004). There is a collection of robust routines for optimization written in the pro-

gramming language Fortran in Madsen et al (2002). A good discussion of numerical optimization with lots of practical advice can be found in the book Gill et al (1981), even though it might be a little old.

The derivation of (7.15) is in (Jondrow et al, 1982, 235 and Kumbhakar and Lovel, 2000, 78). Equation (7.16) is from Jondrow et al (1982, 235). The mode estimator for efficiency in (7.18) is from Jondrow et al (1982, 235), and Battese and Coelli (1988, 392) proposed the estimator in (7.21).

7.14 Appendix: Derivation of the log likelihood function

We will now derive the density in Eq. 7.9 on page 211.

Before we do the integration let us work out the product of the densities

$$\begin{aligned}\varphi_v(\epsilon+u)\,\varphi_u(u) &= \frac{1}{\sqrt{2\,\pi\,\sigma_v^2}}\,e^{-\frac{1}{2}\frac{(\epsilon+u)^2}{\sigma_v^2}}\;\frac{2}{\sqrt{2\pi\,\sigma_u^2}}\,e^{-\frac{1}{2}\frac{u^2}{\sigma_u^2}}\\ &= \frac{1}{\pi\,\sqrt{\sigma_u^2\,\sigma_v^2}}e^{-\frac{1}{2}\frac{u^2}{\sigma_u^2}-\frac{1}{2}\frac{(\epsilon+u)^2}{\sigma_v^2}}\\ &= \frac{1}{\pi\,\sqrt{\sigma_u^2\,\sigma_v^2}}e^{-\frac{1}{2}\frac{\left(\sigma_u^2+\sigma_v^2\right)u^2+2\,\sigma_u^2\,\epsilon\,u+\sigma_u^2\,e^2}{\sigma_u^2\,\sigma_v^2}}.\end{aligned}$$

Now, we can do the integration. This involves some lengthy and nasty steps,so either skip it or sharpen your pen:

$$\begin{aligned}\varphi_\epsilon(\epsilon) &= \int_0^\infty \varphi_v(\epsilon+u)\varphi_u(u)\,du\\ &= \frac{1}{\pi\,\sqrt{\sigma_u^2\,\sigma_v^2}}\int_0^\infty e^{-\frac{1}{2}\frac{\left(\sigma_u^2+\sigma_v^2\right)u^2+2\,\sigma_u^2\,\epsilon\,u+\sigma_u^2\,\epsilon^2}{\sigma_u^2\,\sigma_v^2}}\,du\end{aligned}$$

(Use a math formula collection to get next)

$$\begin{aligned}&= \frac{1}{\pi\,\sqrt{\sigma_u^2\,\sigma_v^2}}\sqrt{\frac{\pi}{2}}\frac{1}{\sqrt{\frac{1}{\sigma_u^2}+\frac{1}{\sigma_v^2}}}\left(1-\operatorname{erf}\left(\frac{\epsilon}{\sqrt{2}\,\sqrt{\frac{1}{\sigma_u^2}+\frac{1}{\sigma_v^2}}\,\sigma_v^2}\right)\right)e^{-\frac{1}{2}\frac{\epsilon^2}{\sigma_u^2+\sigma_v^2}}\\ &= \frac{1}{\sqrt{2\,\pi}\,\sqrt{\sigma_v^2+\sigma_u^2}}\left(1-\operatorname{erf}\left(\frac{\epsilon}{\sqrt{2}\,\sqrt{\sigma_v^2+\sigma_u^2}}\sqrt{\frac{\sigma_u^2}{\sigma_v^2}}\right)\right)e^{-\frac{1}{2}\frac{\epsilon^2}{\left(\sigma_u^2+\sigma_v^2\right)}}\end{aligned} \tag{7.23}$$

(Set $\sigma^2=\sigma_v^2+\sigma_u^2$ and $\lambda=\sqrt{\frac{\sigma_u^2}{\sigma_v^2}}$)

$$= \frac{1}{\sqrt{2\,\pi\sigma^2}} \left(1 - \text{erf}\left(\frac{\epsilon}{\sqrt{2}\,\sqrt{\sigma^2}}\,\lambda\right)\right) e^{-\frac{1}{2}\frac{\epsilon^2}{\sigma^2}}$$

$$= \frac{1}{\sqrt{2\,\pi\sigma^2}} \left(1 + \text{erf}\left(-\frac{\lambda\epsilon}{\sqrt{2\sigma^2}}\right)\right) e^{-\frac{1}{2}\frac{\epsilon^2}{\sigma^2}} \qquad (\text{erf}(x) = \tfrac{2}{\sqrt{\pi}} \textstyle\int_0^x e^{-x^2}\,dx)$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}}\, 2\,\Phi\left(-\frac{\lambda\epsilon}{\sqrt{\sigma^2}}\right)\, e^{-\frac{1}{2}\frac{\epsilon^2}{\sigma^2}} \qquad (\Phi \text{ is the normal distribution})$$

$$= \frac{\sqrt{2}}{\sqrt{\pi\sigma^2}}\,\Phi\left(-\frac{\lambda\epsilon}{\sqrt{\sigma^2}}\right)\, e^{-\frac{1}{2}\frac{\epsilon^2}{\sigma^2}} \tag{7.24}$$

where $\sigma^2 = \sigma_v^2 + \sigma_u^2$ and $\lambda = \sqrt{\frac{\sigma_u^2}{\sigma_v^2}}$.

The *error function* $\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\,dt$ has the following property: $\text{erf}(-x) = -\text{erf}(x)$. Its relationship with the normal distribution is given by $\Phi(x) - \frac{1}{2} = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{1}{2}t^2}\,dt = \frac{1}{2}\,\text{erf}\left(\frac{x}{\sqrt{2}}\right)$ such that $\Phi(x) = \frac{1}{2}\left(1 + \text{erf}\left(\frac{x}{\sqrt{2}}\right)\right)$.

The log of this density is

$$\log \varphi_\epsilon(\epsilon) = -\tfrac{1}{2}\log\left(\frac{\pi}{2}\right) - \tfrac{1}{2}\log\sigma^2 + \log\Phi\left(-\frac{\epsilon\lambda}{\sqrt{\sigma^2}}\right) - \tfrac{1}{2}\frac{\epsilon^2}{\sigma^2}. \tag{7.25}$$

Chapter 8
Additional Topics in SFA

8.1 Introduction

In this chapter, we continue our coverage of Stochastic Frontier Analysis (SFA). We extend the use of SFA to the estimation of general multi-input, multi-output productions functions and show how to estimate cost functions rather than the production functions that we focused on in Chap. 7. We also discuss hypothesis testing within an SFA framework. We conclude with a more methodological discussion of possible problems related to the use of SFA.

8.2 Stochastic distance function models

One limitation of standard SFA models is that they only allow for the analysis of production functions: i.e. situations with one output. We would ideally also be able to model situations with multiple inputs and outputs. There are two possible solutions to this problem.

One is to use cost functions, as we will show in section 8.4 on page 244. However, these require other types of data, namely information on costs, prices and output quantities instead of input and output quantities. We return to this approach in Sect. 8.4 below.

Another option is to use distance or efficiency functions directly on the usual data set: i.e. when we have data for multiple inputs and multiple outputs and no prices. We begin with this approach.

We have defined Farrell input E and output F efficiency on page 26 as

$$E(x,y) = \min\{ E > 0 \mid (Ex, y) \in T \}$$
$$F(x,y) = \max\{ F > 0 \mid (x, Fy) \in T \}.$$

P. Bogetoft and L. Otto, *Benchmarking with DEA, SFA, and R*, International Series in Operations Research & Management Science 157, DOI 10.1007/978-1-4419-7961-2_8,

When we want to parameterize, i.e. use a specific function with parameters for $E(x, y)$ and $F(x, y)$, it is easier to use the inverse distance, the Shephard input D_i and output D_o distance functions

$$D_i(x, y) = \max\{\, D > 0 \mid \left(\frac{x}{D}, y\right) \in T \,\} = \frac{1}{E(x, y)}$$
$$D_o(x, y) = \min\{\, D > 0 \mid \left(x, \frac{y}{D}\right) \in T \,\} = \frac{1}{F(x, y)}.$$

Distance functions can be thought of in a normative manner as a measure of performance. Indeed, this is how we have used the concept throughout most of this book. However, distance functions can also be considered a descriptive device that indicates one of several equivalent ways to describe a technology. In this section, as in Chap. 3, we mainly take the descriptive approach. We will model the technology via a distance function. We can then derive the technology set

$$T = \{\, (x, y) \in \mathbb{R}^m_+ \times \mathbb{R}^n_+ \mid D_i(x, y) \geq 1 \,\}$$
$$T = \{\, (x, y) \in \mathbb{R}^m_+ \times \mathbb{R}^n_+ \mid D_o(x, y) \leq 1 \,\}.$$

if we need it. In many cases, however, we do not need to know T because we use it only to gauge the performance of a given firm, which we can do equally well directly via D_i(x,y); the lower D_i(x,y) is, the better (x, y) is performing.

Not all functions are distance functions and therefore can be interpreted as describing a technology. Thus, in our estimations, we must restrict the types of functions we estimate from our data.

Consider an input distance $D_i(x, y)$. For a fully efficient firm, we have $D_i(x, y) = 1$ on the boundary of T. For an inefficient firm, we have $D_i(x, y) > 1$ corresponding to the interior of T.

Also, $D_i(x, y)$ is homogeneous of degree 1 in x, as can be seen from the following computations

$$\begin{aligned} D_i(tx, y) &= \max_\theta\{\, \theta \mid \left(t\frac{x}{\theta}, y\right) \in T \,\} \\ &= \max_\lambda\{\, \lambda t \mid (\frac{x}{\lambda}, y) \in T \,\} \qquad \left(\frac{\theta}{t} = \lambda\right) \\ &= t \max_\lambda\{\, \lambda \mid (\frac{x}{\lambda}, y) \in T \,\} \\ &= tD_i(x, y). \end{aligned}$$

At this point, we do not need to know any further properties of $D_i(x, y)$.

Let us now introduce a variable $u \geq 0$ such that

$$D_i(x, y) = e^u$$

It follows that $D_i(x, y) = e^u = 1$ and $D_i(x, y) = e^u > 1$ when $u = 0$ and $u > 0$. We can therefore interpret u as a measure of inefficiency. Taking logs, we

can rewrite this as

$$\log(D_i(x, y)) = u$$

We therefore get

$$\log E = \log \frac{1}{D_i} = -\log D_i = -u$$

and

$$E = e^{-u}.$$

Due to this homogeneity, we have

$$x_m D_i\Big(\frac{x}{x_m}, y\Big) = D_i(x, y)$$

such that by taking the log, we get

$$\log x_m + \log D_i\Big(\frac{x}{x_m}, y\Big) = \log D_i(x, y)$$

or

$$\begin{aligned} -\log x_m &= \log D_i\Big(\frac{x}{x_m}, y\Big) - \log D_i(x, y) \\ &= \log D_i\Big(\frac{x}{x_m}, y\Big) - u. \end{aligned}$$

We can turn this into a stochastic model by adding a random error v to get

$$-\log x_m = \log D_i\Big(\frac{x}{x_m}, y\Big) + v - u$$

or

$$\log\Big(\frac{1}{x_m}\Big) = \log\Big(D_i\Big(\frac{x}{x_m}, y\Big)\Big) + v - u. \tag{8.1}$$

We assume that the terms v and u are independent and normally distributed, $v \sim N(0, \sigma_v^2)$ and $u \sim N_+(0, \sigma_u^2)$, where u is only half–normal to ensure $u \geq 0$.

Note that Eq. (8.1) uses precisely the same form as the stochastic production frontier model from Chap. 7. We can therefore use the methods herein to estimate this equation and estimate a model of a general multi-input, multi-output production structure. If we insert a more parametric functional form for the distance function, D_i, we have an *estimable stochastic distance function*.

Numerical example in R: Single-output milk producers

Consider again the milk data that we used for the SFA production function in Sect. 7.9 on page 215, where we estimated a simple model for milk production; the output was kg milk, and the inputs were veterinary expenses, energy and the number of cows.

Next, we use the same data to estimate an input distance function in the simple functional form of a Cobb-Douglass type function: i.e. a linear function using the log of the variables. (Other functional forms are introduced later.) The estimable function in Eq. (8.1) now becomes the estimable distance function for milk production

$$\log \frac{1}{\text{cows}} = \alpha_0 + \alpha_1 \log\Big(\frac{\text{energy}}{\text{cows}}\Big) + \alpha_2 \log\Big(\frac{\text{vet}}{\text{cows}}\Big) + \alpha_3 \log(\text{milk}) + v - u\,.$$

The SFA production function is

$$\log(\text{milk}) = \beta_0 + \beta_1 \log(\text{vet}) + \beta_2 \log(\text{energy}) + \beta_3 \log(\text{cows}) + v - u.$$

We estimate these two SFA function, the distance function and the production function, using with the following lines of code

```
library(Benchmarking)
milkData <- read.csv( "milkProd.csv" )
x <- with(milkData, cbind(vet, energy, cows))
y <- matrix(milkData$milk)
# Production function SFA
milkProd <- sfa(log(x), log(y))
teProd <- te.sfa(milkProd)
# Distance function SFA
Y <- 1/milkData$cows
X <- with(milkData, cbind("vet"=vet/cows, energy=energy/cows,
         milk=milk))
milkDist <- sfa(log(X),log(Y))
teDist <- te.sfa(milkDist)
```

The correlation between the efficiencies estimated using the distance function approach, `teDist` and the former production function approach `teProd` is illustrated in Fig. 8.1. The difference between the production function and the distance function approach is small, and this is no surprise given that the figures are estimated for the same firms and the same technology in two different ways.

Numerical example in R: Multi-output pig producers

The previous example did not really capitalize on the potential of the distance function approach as compared to the production function because it also had only one output. We now consider an example with 248 firms, pig producers that produce 2 outputs and using 6 inputs. The outputs are pig (y_4) and crop (y_2) production, and the inputs are fertilizer (x_1), feed (x_2), land (x_3), labor (x_4), machinery (x_5), and other capital (x_6). In raising pigs, most farmers also produce crops to feed the pigs. Labor and capital are used not just directly for pig-raising but also on the field. The data are in the csv file "pigdata.csv".

```
library(Benchmarking)
d <- read.csv("pigdata.csv")
# SFA distance function
```

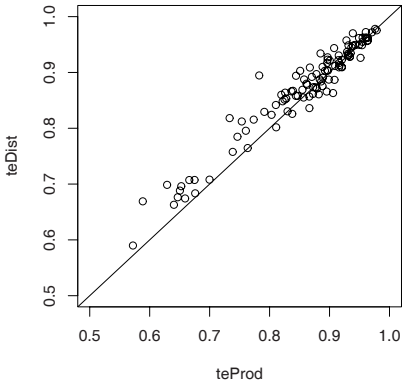


Fig. 8.1 Efficiencies of milk producers based on production (teProd) and distance (teDist) functions

```
X <- with(d,cbind(x2=x2/x1,x3=x3/x1,x4=x4/x1,x5=x5/x1,
                  x6=x6/x1,y2,y4))
Y <- matrix(d$x1,ncol=1)
dist <- sfa(log(X), -log(Y))
te <- te.sfa(dist)
# DEA
x <- with(d,cbind(x1,x2,x3,x4,x5,x6))
y <- with(d,cbind(y2,y4))
Dea <- dea(x,y,"ORIENTATION"="in")
E <- eff(Dea)
Fea <- dea(x,y,"ORIENTATION"="out")
FF <- eff(Fea)
```

Graphs of SFA distance efficiencies are shown in Fig. 8.2.

In Fig. 8.3, we compare the efficiency figures achieved using the distance function approach to those obtained based on DEA input efficiency for the same inputs and outputs. The figure displays a pattern that we have seen before (Fig. 7.12 on page 224); the range and spread of the DEA efficiencies are larger than the range and spread of the distance function SFA efficiencies.

In these calculations, we have used x_1 as the special input variable that we select for the norming and left-hand side of the distance function expression. If we instead used x_6 for this purpose, we would get essentially the same estimates of the model and the individual efficiencies; i.e. it does not matter which input we use as the y-variable in the estimation.

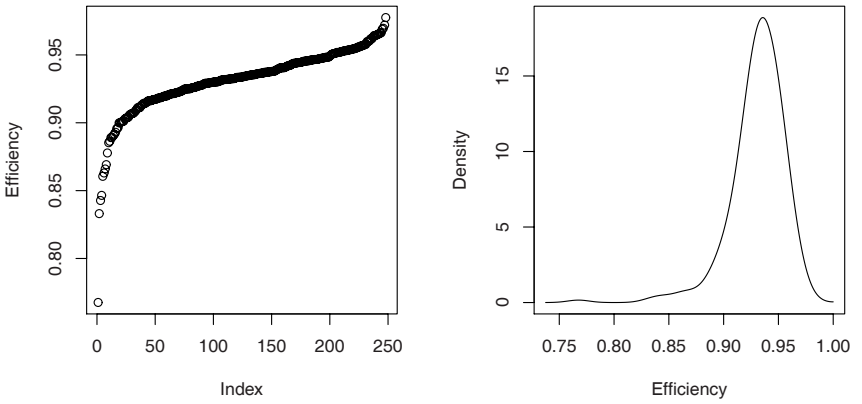


Fig. 8.2 Pig producer efficiencies based on SFA distance function—sorted and as distribution density

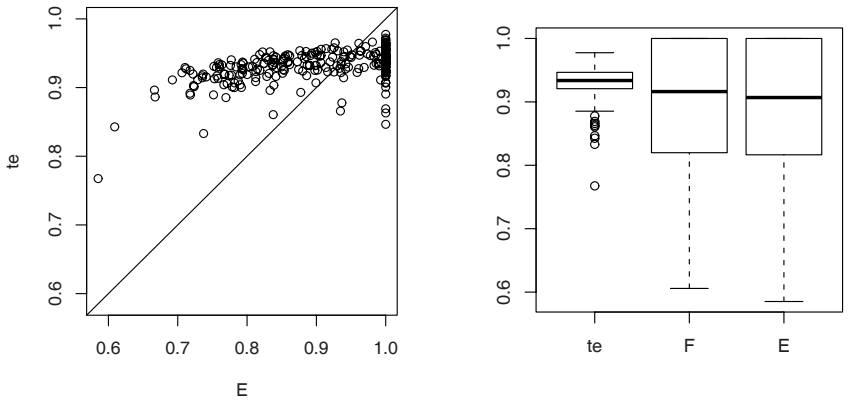


Fig. 8.3 Comparison of pig producer efficiencies according to SFA distance function (te), reciprocal Farrell output measure (F), and Farrell input measure (E)

8.2.1 Estimating an output distance function

We can repeat almost everything we have said about the input efficiency function for the output efficiency function; the only difference is that we now divide all outputs by one output and use that output as the explanatory variable, the variable on the left-hand side.

The output distance function is homogeneous in terms of output

$$\begin{aligned} D_o(x,ty) &= \min_{\phi}\{\,\phi > 0 \mid \big(x, \frac{ty}{\phi}\big) \in T\,\} \\ &= \min_{\psi}\{\,t\psi \mid \big(x, \frac{y}{\psi}\big) \in T\,\} \qquad \big(\frac{\phi}{t} = \psi\big) \\ &= tD_o(x,y) \end{aligned}$$

such that $y_n D_o\big(x, \frac{y}{y_n}\big) = D_o(x,y)$ and $\log y_n + \log\big(D_o(x, \frac{y}{y_n})\big) = \log\big(D_o(x,y)\big)$. The resulting estimation equation can be written in the same way as the input distance function in Eq. (8.1) except that now the outputs are normalized with respect to one of the outputs, whereas the inputs are left unchanged.

The estimable equation then looks like

$$\log y_n = -\log D_o\big(x, \frac{y_n}{y}\big) + v - u$$

where the interpretation of $\frac{y_n}{y}$ is $\big(\frac{y_n}{y_1}, \ldots, \frac{y_n}{y_{n-1}}, 1\big)$.

8.3 Functional forms

Thus far, we have considered only parametric functions of the Cobb-Douglas functional form: i.e. linear or log-linear functional forms. However, stochastic frontier analysis can be used with many other functional forms, and the `sfa` function of the benchmarking package can estimate any functional form that can be made linear in its parameters β; it can also be used with many non-linear models. For instance, the function can be linear, linear in logarithms (Cobb-Douglas), quadratic, quadratic in logarithms (translog), or contain any higher-order exponents (cubic, ...). Furthermore, all of these functions can be estimated with partly log- and partly non-log transformed variables and with other transformations of the variables (e.g. square roots, exponentials).

In this section, we present a brief overview of the possible approximations that are linear in their parameters and that have shown to be useful in empirical applications.

8.3.1 Approximation of functions

There are many methods for approximation of functions. Let us consider a simple class of approximations known as Taylor expansions.

Consider a function $f(z)$ in which z is a number or a m–vector. The function f can be a production function, in which case $z = x$ represents the inputs, or it can be a cost function, in which case $z = (p, y)$ represents input prices and outputs; it can also be a distance function, in which case $z = (x, y)$ represents inputs and outputs.

The simplest approximation is approximation using a constant

$$f(z) = a_0.$$

A more interesting approximation is the linear form

$$f(z) = a_0 + az,$$

and if z is a vector, then az is a vector product (inner product), and we can write it as

$$f(z) = a_0 + \sum_{i=1}^{m} a_i z_i = a_0 + a_1 z_1 + \cdots + a_m z_m.$$

A Cobb-Douglas function $y = Az_1^{a_1} \ldots z_m^{a_m}$ is a linear function if we take the log such that $\log y = a_0 + a_1 \log z_1 + \cdots + a_m \log z_m$ where $a_0 = \log A$.

The linear function has its drawbacks; all first-order derivatives are constants, and all second-order derivatives are zero.

An approximation without these drawbacks is the quadratic approximation, the second-order Taylor expansion,

$$f(z) = a_0 + az + \tfrac{1}{2} Bz^2.$$

When z is a m–vector, then B is a matrix, and

$$\begin{aligned} f(z) = f(z_1, \ldots, z_m) &= a_0 + az + \tfrac{1}{2} z' Bz \\ &= a_0 + \sum_{i=1}^{m} a_i z_i + \tfrac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} z_i B_{ij} z_j \\ &= a_0 + \sum_{i=1}^{m} a_i z_i + \tfrac{1}{2} \sum_{i=1}^{m} \sum_{j=i}^{m} (B_{ij} + B_{ji}) z_i z_j. \end{aligned}$$

Often, the B matrix is considered to be symmetric. B is the second-order derivative w.r.t. z. The order of differentiation makes no difference, and the approximation becomes

$$f(z_1, \ldots, z_m) = a_0 + \sum_{i=1}^{m} a_i z_i + \sum_{i=1}^{m} \sum_{j=i}^{m} B_{ij} z_i z_j.$$

The first-order derivative in this case is

$$\begin{aligned} \frac{\partial f(z)}{\partial z_h} &= a_h + \sum_{j=1}^{m} B_{hj} z_j \\ &= a_h + B_h.z \end{aligned}$$

and as a vector first-order derivative, it is

$$\frac{\partial f(z)}{\partial z} = a + Bz.$$

Note that now, the first-order derivative is not a constant; instead, it is a linear function.

We often use the log of variables such that we consider

$$\log(f(z)) = a_0 + \sum_{i=1}^{m} a_i \log z_i + \tfrac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} B_{ij} \log z_i \log z_j. \tag{8.2}$$

This function form, quadratic in logs, is called a *translog function*. When $f(z)$ is a cost function, the translog function has some very nice properties that make it very useful for empirical analysis, and it is one of the most used functional forms in applied economics. We will return to this concept.

8.3.2 Homogeneous functions

A function is homogeneous if $f(tz) = tf(z)$ for $t > 0$. When we require the function to be homogenous, we cannot use all of the above approximations.

The constant approximation is not usable because

$$f(z) = a_0 \Rightarrow f(tz) = a_0 \neq tf(z) = ta_0,$$

and the linear approximation

$$f(tz) = a_0 + a(tz) = a_0 + taz \neq t(a_0 + az)$$

is only usable if $a_0 = 0$. Thus, the linear approximation is only homogeneous for $a_0 = 0$.

For the quadratic approximation, we find

$$f(tz) = a_0 + taz + \tfrac{1}{2}(tz)'B(tz) = a_0 + taz + \tfrac{1}{2}t^2 z'Bz.$$

This form is not homogeneous at all.

The homogeneity of log linear functions is equivalent to

$$\log(f(tz)) = \log(tf(z)) = \log f(z) + \log t. \tag{8.3}$$

The log linear approximation is

$$\begin{aligned}\log(f(tz)) &= a_0 + \sum_{i=1}^{m} a_i \log(tz_i) \\ &= a_0 + \sum_{i=1}^{m} a_i (\log z_i + \log t) \\ &= a_0 + \sum_{i=1}^{m} a_i \log z_i + \Big(\sum_{i=1}^{m} a_i\Big) \log t \\ &= \log f(z) + \Big(\sum_{i=1}^{m} a_i\Big) \log t.\end{aligned}$$

We can see from Eq. (8.3) that the log linear function is homogeneous when $\sum_{i=1}^{m} a_i = 1$.

The translog function is according to Eq. (8.3) homogeneoushomogeneity!translog when the following equation is equal to translog function plus $\log t$:

$$\begin{aligned}\log\big(f(tz)\big) = a_0 &+ \sum_{i=1}^{m} a_i \log z_i + \Big(\sum_{i=1}^{m} a_i\Big) \log t \\ &+ \tfrac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} B_{ij} (\log z_i + \log t)(\log z_j + \log t).\end{aligned}$$

If we multiply the product in the last term, the equation emerges as equal to

$$\begin{aligned}a_0 + &\sum_{i=1}^{m} \log z_i + \Big(\sum_{i=1}^{m} a_i\Big) \log t \\ &+ \tfrac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} B_{ij} \log z_i \log z_j \\ &+ \tfrac{1}{2} \sum_{i=1}^{m} \Big(\sum_{j=1}^{m} B_{ij}\Big) \log z_i \log t \\ &+ \tfrac{1}{2} \sum_{j=1}^{m} \Big(\sum_{i=1}^{m} B_{ij}\Big) \log z_j \log t \\ &+ \tfrac{1}{2} \sum_{i=1}^{m} \Big(\sum_{j=1}^{m} B_{ij}\Big) (\log t)^2.\end{aligned}$$

For this equation to be equal to the translog function plus $\log t$, the last three terms should be zero, and the term $\big(\sum_{i=1}^{m} a_i\big) \log t$ should be equal to $\log t$. This is clearly the case when

$$\sum_{i=1}^{m} a_i = 1 \tag{8.4}$$

$$\sum_{i=1}^{m} B_{ij} = 0, \qquad j = 1, \ldots, m, \tag{8.5}$$

$$\sum_{j=1}^{m} B_{ij} = 0, \qquad i = 1, \ldots, m. \tag{8.6}$$

Note that if B is symmetrical, $B_{ij} = B_{ji}$, and then the last condition is superfluous.

A nice feature of the translog function is that when we estimate the parameters B, we estimate the second-order derivative that determines how the inputs and outputs interact. In this way, we let the data determine the latter, i.e. whether the inputs and outputs are substitutes or complements.

8.3.3 The translog distance function

We can now use a translog function to approximate $\log\big(D_i\big(\frac{x}{x_m}, y\big)\big)$. In the following translog function, we omit the last input x_m; it would have been $\frac{x_m}{x_m} = 1$ and $\log 1 = 0$,

$$\begin{aligned}\log\Big(\frac{1}{x_m}\Big) =& a_0 + \sum_{i=1}^{m-1} a_i \log\frac{x_i}{x_m} + \sum_{j=1}^{n} b_j \log y_j \\ &+ \tfrac{1}{2}\sum_{i=1}^{m-1}\sum_{j=1}^{m-1} A_{ij} \log\frac{x_i}{x_m} \log\frac{x_j}{x_m} \\ &+ \tfrac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} B_{ij} \log y_i \log y_j \\ &+ \tfrac{1}{2}\sum_{i=1}^{m-1}\sum_{j=1}^{n} C_{ij} \log\frac{x_i}{x_m} \log y_j + v - u.\end{aligned} \tag{8.7}$$

The restrictions on the translog parameters from page 242 have many implications. Equation (8.4) implies that we only need to estimate $a_1, \ldots, a_{m-1}$ and then use $a_m = 1 - \sum_{i=1}^{m-1} a_i$ such that the sum restriction is automatically fulfilled. In the same way, the other restrictions are also automatically fulfilled. It is not really a surprise that the restrictions for a homogeneous translog are fulfilled because we choose the estimating equation in (8.7) to make it homogenous by dividing all inputs by input m, x_m. When we are using maximum likelihood estimation, it does not matter which parameter is estimated as a residual because the maximum likelihood estimator is unique.

A translog approximation of the output distance function looks like this:

$$\begin{aligned}\log\left(\frac{1}{y_n}\right) =& a_0 + \sum_{i=1}^{m} a_i \log x_i + \sum_{j=1}^{n-1} b_j \log \frac{y_j}{y_n} \\ &+ \tfrac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} A_{ij} \log x_i \log x_j \\ &+ \tfrac{1}{2}\sum_{i=1}^{n-1}\sum_{j=1}^{n-1} B_{ij} \log \frac{y_i}{y_n} \log \frac{y_j}{y_n} \\ &+ \tfrac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n-1} C_{ij} \log x_i \log \frac{y_j}{y_n} \\ &+ v - u.\end{aligned}$$

Translog can also be used for production functions and cost functions.

8.4 Stochastic cost function

Having previously considered the production function, we will now address the cost function.

Let us begin with the case of a *single-input, multi-output technology*; i.e. x is one-dimensional and y is n-dimensional. In this case, inefficiency will lead to excessive values of x, and we would therefore ideally like to estimate a stochastic frontier

$$x = c(y;\alpha) + v + u$$

where v is the symmetric noise term and u is the positive inefficiency term; e.g., $v \sim N(0,\sigma_v^2)$ and $u \sim N_+(0,\sigma_u^2)$. Note that this is almost equivalent to the case of a SFA production function, and we would therefore ideally want to estimate the stochastic frontier function

$$y = g(x;\beta) + v - u$$

where y is one-dimensional and x is m-dimensional. The only real difference is that we add the inefficiency figure in a single-input (cost) setting, whereas we subtract it in a single-output (production) setting. This difference reflects how inefficiency increases the use of inputs or reduces the amount of output.

Now, we can rewrite the single-input equation as

$$-x = -c(y;\alpha) - v - u$$

The symmetric term is now $-v$ but has the same distribution as v. The asymmetric term is now $-u$, as in the production function equation. Hence, we can estimate single-input stochastic frontiers in the same way as we estimate single-output stochastic production function; we simply use the input with an opposite sign as the dependent variable and the functional form also with an opposite sign as the independent part.

Note also that in such cases, we are effectively measuring cost efficiency and presuming that all firms face the same input prices. To be cost efficient, firms must be both technically efficient and allocatively efficient, as explained in Chap. 2.

Single-input cost functions are quite comment in applications. It is often the case that inputs are aggregated into total costs, whereas outputs are specified in more detail and cannot be aggregated because we lack prices on the output side. For example, this is the case in many public sector applications in which there are no market prices associated with the services produced.

Another common situation arises when we know the services y, the price of inputs w, and the total costs of production $c = wx$, but lack information about the different physical inputs used. In such cases, we can instead estimate a general cost function. A *general cost function* or simply a cost function explains total cost as a function of input prices and output quantities. The function $c()$ above is a special case corresponding to just one type of input (total costs) and therefore also only one input price, 1.

An advantage of general cost functions as opposed to production functions is that they can easily handle more than one output. All questions that can be answered by a production function can also be answered by a cost function, often more easily by using Shephard's lemma directly or indirectly. However, the cost functions require prices. In accounting, and in many other kind of data sources, quantities and value are more often available than prices, and this makes the cost function less attractive than the production function. However, much accounting data is in the form of value figures, i.e. prices multiplied with quantities. The production function approach is therefore also in need for prices to recapture the quantities, and in this way, data requirement are the same whether we estimate production or cost functions based on most types of accounting data—data availability rarely determines the mode of analysis. If we only have accounting data, we can assume that all firms have the same prices and the use the values from the accounts as quantities—or perhaps use a price index from a stastistical bureau to deflate values to quantities. In this case, it does not make sense to estimate a cost function because we assume that there is no variation in input prices between the firms. Neither does it make sende to assume that all firms use the same quantities and then calculate firm-specific prices.

The cost functions shows the minimum cost of producing the output combination y when the input prices w and the technology set T are given:

$$c(w, y) = \min_x \{\, wx \mid (x, y) \in T \,\}.$$

Therefore, the actual or observed cost is greater than or equal to the minimum cost

$$wx \geq c(w, y) \qquad \text{for all } (x, y) \in T \tag{8.8}$$

Cost efficiency (i.e., minimal cost compared to actual costs)

$$CE = \frac{c(w, y)}{wx}.$$

is therefore at the most 1; $CE \leq 1$, for $(x, y) \in T$. As before, we may parameterize cost efficiency via an inefficiency term u; i.e., we may let

$$CE = e^{-u}.$$

where $u \leq 0$ when $CE \leq 1$, for $(x, y) \in T$

We introduce a multiplicative error term v such that cost efficiency becomes

$$CE = \frac{c(w, y)\, e^{v}}{wx}$$

and therefore,

$$wx = \frac{c(w, y)e^{v}}{CE} = \frac{c(w, y)\, e^{v}}{e^{-u}} = c(w, y)\, e^{v} e^{u}.$$

Estimations of CE can be based on this equation, which is just like the one that we have been using in our stochastic frontier analysis. If we take the log, we find a familiar-looking equation

$$\log(wx) = \log\big(c(w, y)\big) + v + u$$

Only the sign of u is different.

We illustrate the method based on a *Cobb-Douglas cost function* where we let $c = \sum_{i=1}^{m} w_i x_i$ such that

$$c = \beta_0\, w_1^{\beta_1} \cdots w_n^{\beta_n}\, e^{\epsilon}$$

where $\epsilon = v + u$, and $\sum_{i=1}^{n} \beta_i = 1$ to ensure that the cost function is homogeneous in the input prices. In logarithmic form, the Cobb-Douglas function is

$$\begin{aligned} \log c &= \beta_0 + \beta_1 \log w_1 + \cdots + \beta_n \log w_n + \epsilon \\ &= \beta_0 + \beta_1 \log w_1 + \cdots + \beta_n \log w_n + v + u \end{aligned} \tag{8.9}$$

where v is an ordinary error term and u is a non-negative term that reflects the inefficiency level.

Cost efficiency can now be measured as

$$CE = e^{-u} = \frac{\beta_0\, w_1^{\beta_1} \ldots w_n^{\beta_n}\, e^{v}}{c}.$$

The density of the error term $\epsilon = v + u$ is asymmetrically skewed because u is non-negative.

Apart from the restrictions on the β's because of the homogeneity of the input prices and the direction of the skewness of the error term, the above Eq. (8.9) corresponds to Eq. (7.5) on page 204. We can thus use the earlier statistical method to estimate the cost function and the level of cost efficiency.

We can see this more directly if we note that v_i is symmetrical. We can rewrite Eq. (8.9) as the statistically equivalent equation

$$-\log c = -\beta_0 - \beta_1 \log w_1 - \cdots - \beta_n \log w_n + v - u.$$

If we just let x denote the vector of minus the logarithm of input prices and let y be minus the logarithm of expenditure, then we can use the same statistical method to estimate the cost function as we used to estimate the production function as described in Eq. (7.5) on page 204.

A simple way to handle *homogeneity in input prices* is to use one of the input prices as a numeraire: i.e. to use

$$-\log\left(\frac{c}{w_n}\right) = -\beta_0 - \beta_1 \log\left(\frac{w_1}{w_n}\right) - \cdots - \beta_{n-1} \log\left(\frac{w_{n-1}}{w_n}\right) + v - u.$$

This is equivalent to

$$-\log c = -\beta_0 - \beta_1 \log w_1 - \cdots - \beta_{n-1} \log w_{n-1} - (1 - \beta_1 + \cdots + \beta_{n-1}) \log w_n + v - u.$$

Therefore, in principle, we do not need a special function in R to estimate a cost function because we already have one for the production function.

Numerical example in R: Pig producers

We continue with the data for pig producers both raising pigs and cultivating crops. Now, however, we use input prices and output as the explanatory variables, the x–variables, and the total cost as the explained variable, the y–variable. We can see this in the following lines of code:

```
> library(Benchmarking)
> d <- read.csv("pigdata.csv")
> W <- with(d, cbind(w1,w2,w3,w5,w6))
> Y <- with(d, cbind(y2,y4))
> costSfa <- sfa(-log(cbind(W,Y)), -log(d$cost))
> summary(costSfa)
            Parameters   Std.err   t-value  Pr(>|t|)
(Intercept)    5.79486   1.56072    3.7129     0.000
xw1           -0.08552   0.43219   -0.1979     0.843
xw2            1.61290   0.79804    2.0211     0.044
xw3            0.10961   0.01751    6.2598     0.000
xw5            0.71253   0.43177    1.6503     0.100
xw6            0.01842   0.21881    0.0842     0.932
```

```
xy2             0.08844   0.01006    8.7956      0.000
xy4             0.82091   0.01858   44.1827      0.000
lambda          1.53937   0.40795    3.7734      0.000
sigma2          0.033593
sigma2v =  0.00996922 ;  sigma2u =  0.02362373
-log likelihood =  -144.533
Convergence =  4
> teCost <- te.sfa(costSfa)
> colnames(teCost) <- "teCost"
```

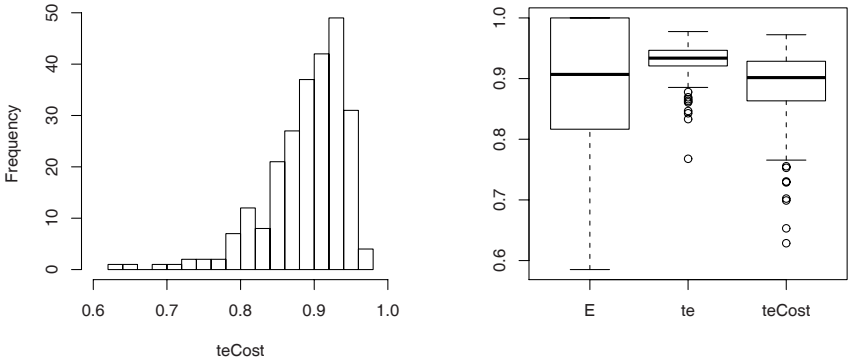


Fig. 8.4 Efficiency for pig procucers estimated from cost function (teCost) compared to DEA (E) and SFA distance function estimates (te) of efficiency

Based on the resulting graph in Fig. 8.4, we see that in this example, the range of efficiency figures is larger when estimated using the cost function than using the distance function. This might be because of the missing value for wages in the data; as a result, labor price is not part of the input prices in the cost function, whereas labor is accounted for directly under the distance function approach. That a larger spread is derived using DEA than using the SFA approach is also demonstrated in Sect. 7.11. Here, as in many other studies, the lack of firm-specific prices complicates the use of the cost function approach.

8.5 Statistical inference

In this chapter and in the previous one, we have estimated the various parameters of stochastic frontier models. Often, we are not just interested in the parameters; rather, we also wish to test hypotheses about the parameters. We will now investigate the various ways in which we can do this depending on the specific form of hypothesis.

8.5.1 Variance of parameters

Many tests take into account the variance of the estimated parameters. Therefore, we will provide a brief description of how to estimate variance as part of the process of optimizing likelihood as described in Sect. 7.6.2. (The material in this subsection is more complicated and may be skipped during a first reading.)

Let f be a density function for a probability distribution such that $\int f\,dy = \int f(y;x,\beta)\,dy = 1$. When we differentiate this equation, we get

$$0 = \int \frac{\partial f}{\partial \beta}\,dy = \int \frac{\partial \log f}{\partial \beta} f\,dy = \int \frac{\partial \ell(\beta)}{\partial \beta} f\,dy = \mathrm{EV}\,\frac{\partial \ell(\beta)}{\partial \beta}$$

where $\ell(\beta) = \log f(y;\beta)$ is the the log-likelihood function. It follows that we have $\mathrm{EV}\,\frac{\partial \ell(\beta)}{\partial \beta} = 0$. Let $\mathrm{Var}\left(\frac{\partial \ell(\beta)}{\partial \beta}\right) = J$ where J is called *Fisher's information matrix*. Based on the central limit theorem, it now follows that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial \log f}{\partial \beta} = \frac{1}{\sqrt{n}} \frac{\partial \ell(\beta)}{\partial \beta}$$

is approximately distributed as $N(0, J)$. By differentiating $\int f\,dy = 1$ twice w.r.t. β, one can also show that $-\mathrm{EV}\,\frac{\partial^2 \ell}{\partial \beta^2} = \mathrm{Var}\,\frac{\partial \ell}{\partial \beta} = J$.

From Eq. (7.6) page 210, we find if we let β^0 be the true value β that

$$\hat{\beta} = \beta - \left(\frac{\partial^2 \ell(\beta)}{\partial \beta^2}\right)^{-1} \frac{\partial \ell(\beta)}{\partial \beta}\,.$$

Based on the above results indicating that $\mathrm{EV}\,\frac{\partial \ell(\beta)}{\partial \beta} = 0$ and $-\mathrm{EV}\,\frac{\partial^2 \ell}{\partial \beta^2} = \mathrm{Var}\,\frac{\partial \ell}{\partial \beta}$, we now derive

$$\mathrm{EV}\,\hat{\beta} = \beta \tag{8.10}$$

and

$$\mathrm{Var}\,\hat{\beta} = \mathrm{Var}\left(-\left(\frac{\partial^2 \ell}{\partial \beta^2}\right)^{-1} \frac{\partial \ell}{\partial \beta}\right) = \left(\mathrm{EV}\,\frac{\partial^2 \ell}{\partial \beta^2}\right)^{-2} \mathrm{Var}\left(\frac{\partial \ell}{\partial \beta}\right) = J^{-2} J = J^{-1}\,. \tag{8.11}$$

Therefore, $\hat{\beta}$ is approximately distributed as $N(\beta, J^{-1})$; i.e.,

$$\hat{\beta} \overset{a}{\sim} N(\beta, J^{-1})\,.$$

An estimate of J can be derived simply by subtracting the empirical mean of $\frac{\partial^2 \ell}{\partial \beta^2}$, which we can derive during the iterative optimization process when we use Newton's methods as we explained in Sect. 7.6.1.

For a straight line, the second-order derivative is zero, and for a curve that is almost straight (i.e., very flat with little curvature), the second-order derivative is very close to zero; in this case, the parameter estimate is made with very little precision. This is illustrated on the left side of Figure 8.5. The right part of the figure show

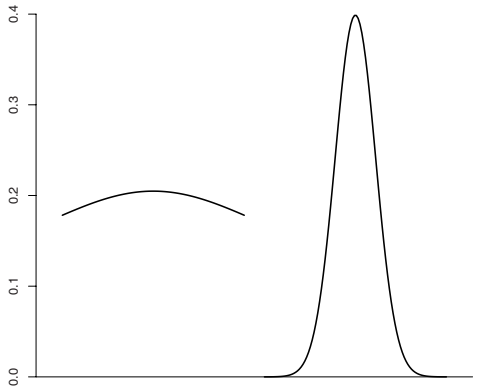


Fig. 8.5 Examples of likelihood functions with small and large curvatures

a curve with high curvature where the parameter is therefore estimated with high precision.

Another way to explain this is to note that as $\hat{\beta} \overset{a}{\sim} N(\beta, J^{-1})$ where J is the Fisher information matrix introduced on page 249 and $J = -\text{EV}\,\frac{\partial^2 \ell}{\partial \beta^2}$, a little curvature in the log-likelihood function corresponds to a small $|J|$, the determinant of the matrix J , and therefore to a large amount of variance J^{-1} of β. Correspondingly a large curvature corresponds to a small amount of variance and a high-precision estimate. Again, this also follows intuitively from the figure.

8.5.2 Hypothesis testing using the t-test

Because $\hat{\beta} \overset{a}{\sim} N(\beta, J^{-1})$, we can use the square root of the diagonal in J^{-1} as the standard errors of the parameters. This is precisely what we do in the `summary` for an `sfa` object. Then we can test all parameters to verify that they are significantly different from 0 using the usual t-test that is also part of the `summary`; this is just like the usual t-test in linear regression except that for linear regressions un-

der standard assumptions, the distributions are the true distributions, whereas the distributions under SFA are only asymptotic distributions.

8.5.3 General likelihood ratio tests

In ordinary regression models, we use the t-test to test whether a variable should be included or not and the F-test to test for a group of variables. In non-standard models like the SFA model, these tests are not always directly available because the models and hypotheses are non-linear in nature.

A hypothesis H_0 can often be formulated in the form of restrictions on β (for instance, indicating that some coordinates are zero or that some of the coordinates sum to zero or one). The alternative H_A is that there are no such restrictions on β:

$$H_0: \text{Restrictions on } \beta$$
$$H_A: \text{No restrictions on } \beta$$

If the hypothesis is true, then the maximum value of the likelihood function $L(\beta)$ does not depend on whether or not we estimate under the alternative hyphotesis. However, if the hypothesis is false and we estimate the model as if it were true, we put severe restrictions on the model, and therefore, the maximum value of the likelihood function will be much smaller. Let L^0 and L^A be the maximum value of the likelihood function under the hypothesis and under the alternative. The ratio

$$Q = \frac{L^0}{L^A} = \frac{\text{Max likelihood with restriction}}{\text{Max likelihood without restriction}}$$

will therefore be close to 1 if the hypothesis is true and far less than 1 if it is false. This test is called the *likelihood ratio test*.

Now

$$2\log\frac{L^0}{L^A} = 2(\log L^0 - \log L^A) = 2(\ell^0 - \ell^A)\,.$$

We now use a Taylor expansion, as we have previously done, to obtain

$$\ell^0 = \ell^A + (\beta^0 - \beta^A)\frac{\partial \ell}{\partial \beta} + \tfrac{1}{2}(\beta^0 - \beta^A)^2\frac{\partial^2 \ell}{\partial \beta^2}\,.$$

Now EV $\frac{\partial \ell}{\partial \beta} = 0$, as we have previously shown, and therefore, the first-order term in the equation is close to zero. This leave us with

$$\ell^0 = \ell^A + \tfrac{1}{2}(\beta^0 - \beta^A)^2\frac{\partial^2 \ell}{\partial \beta^2}$$

or

$$2(\ell^0 - \ell^A) = (\beta^0 - \beta^A)^2 \frac{\partial^2 \ell}{\partial \beta^2}.$$

On the right-hand side, we have $-\frac{\partial^2 \ell}{\partial \beta^2} \simeq -\text{EV}\frac{\partial^2 \ell}{\partial \beta^2} = J$ where J is the Fisher information matrix. Thus, $-2(\ell^0 - \ell^A)$ is distributed almost as $(\beta^0 - \beta^A)^2 J$.

We have already found that $\hat{\beta}$ is distributed almost as $N(\beta, J^{-1})$, and therefore that $\beta^0 - \beta^A$ is distributed almost like $N(0, J^{-1})$. As a result, we have $(\beta^0 - \beta^A) J^{\frac{1}{2}}$ is distributed like $N(0, 1)$,and therefore $(\beta^0 - \beta^A)^2 J$ is distributed like a chi-squared distribution. The number of degrees of freedom is the number of independent restrictions in the hypothesis or the difference between the number of independent parameters in the two hypotheses.

For $Q = \frac{L^0}{L^A}$, we hence have that *under the null-hypothesis* $-2 \log Q$ *is asymptotic distributed like* χ^2.

8.5.4 Is the variation in efficiency significant?

One hypothesis that we may test is if there is actually any inefficiency in a sector when the alternative is that variations in performance simply reflect noise. Specifically, we will test the null hypothesis that $\sigma_u^2 = 0$ against the alternative that $\sigma_u^2 > 0$:

$$H_0 : \sigma_u^2 = 0$$
$$H_A : \sigma_u^2 > 0$$

Toward this end, we can use both a *t*-test and a likelihood ratio test, as we will now illustrate using the data on milk producers from Sect. 7.9 on page 215.

First, we will consider the *t*-test. We estimate the model under the alternative hypothesis in which $\sigma_u^2 > 0$; that is, we estimate the model as an SFA model. We do this with the following commands and corresponding output

```
> library(Benchmarking)
> milkdata <- read.csv("milkProd.csv")
> y <- cbind("milk"=milkdata$milk)
> x <- with(milkdata,cbind(vet,energy,cows))
> m <- sfa(log(x),log(y))
> summary(m)
            Parameters   Std.err   t-value  Pr(>|t|)
(Intercept)    7.52014   0.32197    23.357     0.000
xvet           0.06281   0.02496     2.517     0.013
xenergy        0.12156   0.03676     3.307     0.001
xcows          0.87879   0.06640    13.235     0.000
lambda         3.59708   0.89964     3.998     0.000
sigma2         0.045685
sigma2v =  0.003277507 ;  sigma2u =  0.04240755
-log likelihood =  -67.82555
```

Table 8.1 OLS and SFA parameter estimates

model	(Intercept)	xvet	xenergy	xcows	log likelihood
OLS	7.103	0.096	0.121	0.859	58.81
SFA	7.521	0.063	0.121	0.879	67.83

Note: $-2\log Q = -2(58.8 - 67.8) = 18.0$

The parameter `lambda` is given by $\lambda = \sqrt{\frac{\sigma_u^2}{\sigma_v^2}}$, and our null-hypothesis is therefore equivalent to $\lambda = 0$. We can test this using a t-test. The t-value in the above summary from the SFA estimation for `lambda` is 3.998, far above the critical value of 1.96. Therefore, we reject the null hypothesis and accept that there are differences in efficiency; the variance σ_u^2 is significantly greater than 0.

The same test can be performed as a likelihood ratio test. Here we compare the parameter estimates in an ordinary regression model, the null-hupothesis where there is no difference between the firms in terms of efficiency, with the estimates in a stochastic frontier model, the alternative hyphotesis where there is a difference; we compare OLS with SFA. The estimated parameters for the OLS and SFA models are shown in Table 8.1. Here we see the from the calculations that the test value is $-2\log Q = 18.0$. This figure must be compared with a chi-squared distribution with 1 degree of freedom, and the 95% the critical value can be found to be 3.84. The test value is thus larger than the critical value and we therefore reject the null-hypothesis. The conclusion is there that there are significantly differences in efficiency between firms. This was the same conclusion we found above with a t-test.

Based on the table, the parameter estimates are more or less the same, but the parameter for veterinary expenses is smaller in SFA, and the parameter for cows is larger.

8.6 Test for constant returns to scale

We will consider three different ways to test for constant returns to scale in a Cobb–Douglas production function, taking differences in efficiency into consideration. The methods themselves are not restricted to use with Cobb–Douglas functions or constant returns to scale but instead carry over to a much broader field of applications.

We test for constant returns to scale in the model

$$y = Ax_1^{a_1} \dots x_n^{a_n},$$

which we immediately rewrite as a log linear function

$$\log y = a_0 + a_1 \log x_1 \cdots + a_n \log x_n \tag{8.12}$$

where $a_0 = \log A$. We can now formulate our hypothesis as

$$H_0 : a_1 + \cdots + a_n = 1$$
$$H_A : a_1 + \cdots + a_n \neq 1 \qquad \text{(two sided alternative)}$$

Numerical example in R: Milk producers

To illustrate the tests in practice, we use a simple model for milk production in which the output is kg milk and the inputs are veterinary expenses, energy and the number of cows.

We use the data for milk production that we also used in section 7.9 on page 215 and again in section 8.2 on page 235.

We first estimate an SFA model using the following R code:

```
> library(Benchmarking)
> d <- read.csv("milkProd.csv")
> y <- log(with(d,cbind(milk)))
> x <- log(with(d,cbind(vet, energy, cows)))
> summary(sfa(x,y))
            Parameters   Std.err   t-value  Pr(>|t|)
(Intercept)    7.52014   0.32197    23.357     0.000
xvet           0.06281   0.02496     2.517     0.013
xenergy        0.12156   0.03676     3.307     0.001
xcows          0.87879   0.06640    13.235     0.000
lambda         3.59708   0.89964     3.998     0.000
sigma2         0.045685
sigma2v =  0.003277507 ;  sigma2u =  0.04240755
log likelihood = 67.82555
```

First, we can see that all parameters are significantly different from 0. We find that the sum of the relevant parameters is $0.063 + 0.121 + 0.879 = 1.063$ such that we do not have constant returns in the estimated parameters; the sum is larger than 1. However, the question should be whether the sum is significantly larger than 1.

We consider three ways to test the hypothesis. They feature varying levels of complexity and generality, meaning that the last methods can be used to test almost anything everywhere, whereas the first method is suited to a more restricted hypothesis.

8.6.1 Rewrite the model: t-test

First, we rewrite the model so that we can test the hypothesis via a simple *t*-test directly from the output of the estimation. We divide all variables in Eq. (8.12) by x_n, or we can substract $\log x_n$ from both sides, remembering that to subtract the log of a variable is the same as dividing by the same variable before taking the log, we get

$$\log \frac{y}{x_n} = a_0 + a_1 \log \frac{x_1}{x_n} + \cdots + a_{n-1} \log \frac{x_{n-1}}{x_n} + (+a_1 + \cdots + a_n - 1) \log x_n \, .$$

Based on this revision, we can see that the hypothesis that there are constant returns to scale, testing for $a_1 + \cdots + a_n = 1$, is equal to the hypothesis that the parameter for the variable $\log x_n$ is equal to zero. This hypothesis can be tested using an ordinary t-test. In R, we do this as below, where we also show the results:

```
> y1 <- log(with(d,cbind(milkPerCow=milk/cows)))
> x1 <- log(with(d,cbind(vetPerCow=vet/cows,
+              energyPerCow=energy/cows, cows)))
> summary(sfa(x1,y1))
               Parameters   Std.err   t-value  Pr(>|t|)
(Intercept)       7.51865   0.31656    23.751     0.000
xvetPerCow        0.06284   0.02234     2.812     0.005
xenergyPerCow     0.12165   0.03527     3.449     0.000
xcows             0.06332   0.03120     2.030     0.044
lambda            3.59783   0.93046     3.867     0.000
sigma2          0.045691
sigma2v =  0.003276627 ;  sigma2u =  0.04241401
-log likelihood =  -67.82554
Convergence =  4
```

Note that the equation `vetPerCow=vet/cows` put a label on the results of the calculations. If we compare these results with the original estimates in the previous subsection, we can see that all parameters are the same except for that of `cows`. This is no surprise given that the model is an equivalent form of the original equation. The estimated parameter for `cows`, 0.063 is an estimate of $+a_1 + \cdots + a_n - 1$, which corresponds to the sum of the parameters 1.063 minus 1, which we calculated in Sect. 7.9.

The t-value for the `cows` parameter is 2.03. This value is above the 97.5% quantile (two–sided alternative) in the t distribution with $108 - 4 = 104$ degrees of freedom, 1.98. Thus, we reject the hypothesis of constant returns to scale.

This method of testing a hypothesis is very often easy to use; we can rewrite a model such that the hypothesis can be tested using a t-test for an estimated parameter.

8.6.2 Linear hypothesis

The next method is to formulate the hypothesis as a linear function of the parameters. This is the scenario including constant returns to scale where the hypothesis is that the sum of the parameters is equal to 1,

$$a_1 + \cdots + a_n = (1, \ldots, 1) \cdot (a_1, \ldots, a_n)' = 1.$$

In general, we want to test the hypothesis $g\beta = g_0$ where g is a vector and g_0 a number; in the above, $g = (1, \ldots, 1)$ and $g_0 = 1$. Let V denote the variance matrix of the parameters β; then the variance of $g\beta$ is $\text{Var}(g\beta) = g\,\text{Var}(\beta)g' = gVg'$, and the standard error is $\sqrt{gVg'}$. Under the hypothesis, the mean of $g\beta$ is g_0 and therefore,

$$t = \frac{g\beta - g_0}{\sqrt{gVg'}} \tag{8.13}$$

has a t distribution.

We can carry out the calculations using the following commands in R:

```
g = matrix(c(0,1,1,1,0),nrow=1)
# sum of parameters
g %*% o$par
# Variance of sum of parameters
g %*% o$vcov %*% t(g)
# t-value for parameters sum to 1
(g %*% o$par - 1) / sqrt(g %*% o$vcov %*% t(g))
```

In the first line, we estimate the model and save the results in the object `o`. The vector g is defined as a matrix with 1 row that will indicate the sum of the parameters; the first 0 in g is exclude the intercept and the last 0 to exclude λ (`lambda`) in the sum. The sum of the parameters is found as a matrix product because `%*%` is the inner product and `o$par` is the parameters in the SFA object. The variance of the sum of the parameters, $g\beta$, is calculated according to formula Eq. (8.13). The t-value, not shown, is 2.00. Because this is larger than the critical value of 1.98, the 97.5% quantile, we reject the null-hypothesis of constant returns to scale.

Note that the t-value calculated will be the same as that calculated in the previous subsection where we found the t-value directly from the estimation output itself.

8.6.3 Likelihood ratio test

The most general method of testing a hypothesis is the likelihood ratio test. We showed in Sect. 8.5.3 that we can test a hypothesis by comparing the value of the likelihood function under the hypothesis with the value under the alternative. The ratio between the two values was called Q, and we showed that $-2\log Q$ under the hypothesis has a chi-squared distribution.

With same data as above, `x` and `y` as the general model from Sect. 7.9, and `y2` and `x2` as variables under the hypothesis of constant returns to scale as in Sect. 8.6.1, the R console looks like

```
> y0 <- log(with(d,cbind(milk/cows)))
> x0 <- log(with(d,cbind(vet/cows, energy/cows)))
> o0 <- sfa(x0,y0)
> # Under the alternative
> oA <- sfa(x,y)
> logQ <- (logLik(o0) - logLik(oA))[1]
> -2*logQ
```

```
[1] 4.422103
> 1-pchisq(-2*logQ, oA$df-o0$df)
[1] 0.03547628
> # Critical value
> qchisq(0.95, oA$df-o0$df)
[1] 3.841459
```

The model is estimated under the hypothesis, and the results are saved in the object `o0`; the results from the general model, the alternative, are saved in the object `oA`. The log likelihood ratio is calculated, after which the test statistic $-2 \log Q$ is calculated and shown to be 4.4. The test statistic is compared to a chi-square distribution whose number of degrees of freedom corresponds to the change in the number of parameters. The change in the number of parameters is 1, so the critical value is the 95% quantile in the chi-squared distribution with 1 degree of freedom, and that quantile is 3.84. We, therefore, reject the null hypothesis because the calculated statistic of 4.4 is larger than 3.84.

8.7 Other distributions of technical efficiency

What if the technical efficiencies are not half–normally distributed? When we use the half–normal distribution, we implicitly assume that most firms have an efficiency level near 1 because the mode of the distribution is 0 and the efficiency is $e^0 = 1$. What if the efficiency of most firms is below 1 and only a few firms have an efficiency level of 1 or near 1? If this is the case, we must consider another distribution of efficiencies. We will look for a continuous distribution with support on the positive axis and mode (peak) away from zero. There are many distributions with this characteristic.

Truncated normal

We have used the half-normal distribution for efficiency. One could also use a truncated normal distribution with an unknown point of truncation. If we use the latter strategy, the SFA model becomes

$$\begin{aligned} y^k &= f(x^k;\beta) + v^k - u^k, \\ & v^k \sim N(0,\sigma_v^2),\ u^k \sim N_+(\mu,\sigma_u^2), \qquad k = 1,\ldots,K. \end{aligned}$$

This formulation does have a nice feature: the mode of the efficiency distribution does not have to be 0, and therefore, the mode of efficiency is not necessarily 1 but can be below 1. The cost of this feature is an extra parameter, the point of truncation μ, which might inflate the standard errors of other parameters because the curvature of the likelihood function is flattened and the determinant of the Fisher

matrix increases, cf. Sect. 8.5.1. Also, this model often results in problems with the convergence of the iterative estimation process.

Exponential

The exponential distribution has been suggested, but its mode is 0. Therefore, it does have the same central characteristics as the normal distribution. The density of the exponential distribution with positive scale parameter β is

$$\varphi(u) = \frac{1}{\beta} e^{-u/\beta}, \qquad (u \geq 0).$$

The exponential distribution is a special instance of the Γ-distribution (the gamma-distribution).

The density of the exponential distribution is shown as the curve marked 1 in Fig. 8.6. Even though the curvature is a little different, the exponential function looks very similar to the half–normal distribution we used in Chap. 7, cf. Fig. 7.5 on page 206. Thus, the implications of using the exponential distribution for efficiencies instead of the half–normal distribution are not noticeable in empirical applications. This conclusion is consistent with the results of related empirical studies.

Gamma

The Γ-distribution with shape parameter λ has the density

$$\varphi(u) = \frac{1}{\Gamma(\lambda)} u^{\lambda-1} e^{-u}$$

where the function Γ is defined by Eulers second integral

$$\Gamma(\lambda) = \int_0^\infty x^{\lambda-1} e^{-x}\, dx, \qquad (\lambda > 0)\,.$$

The Γ–function fulfills the functional equation

$$\Gamma(\lambda) = \lambda\Gamma(\lambda), \qquad (\lambda > 0)\,.$$

The parameter $f = 2\lambda$ is called the degrees of freedom. The distribution have mode in 0 for $0 < \lambda \leq 1$ and in $\lambda - 1$ for $1 < \lambda$. One can see that the exponential distribution is a Γ-distribution with shape parameter $\lambda = 1$.

The Γ-distribution with shape parameter λ and scale parameter $\beta >$ has the density

$$\varphi(u) = \frac{1}{\beta^\lambda \Gamma(\lambda)} u^{\lambda-1} e^{-u/\beta}.$$

The mean is $\beta\lambda$, and the variance is $\beta^2\lambda$. The Γ-distribution is a rather flexible distribution. For small values of the shape parameter λ, the density has a thick tale to the right, and for large values of the shape parameter, the distribution looks like the normal distribution with parameters $(\lambda\beta, \lambda\beta^2)$.

A series of densities for the Γ-distribution is shown in Fig. 8.6. Note that the curve of the shape parameter $\lambda = 1$ corresponds to the density of the exponential distribution.

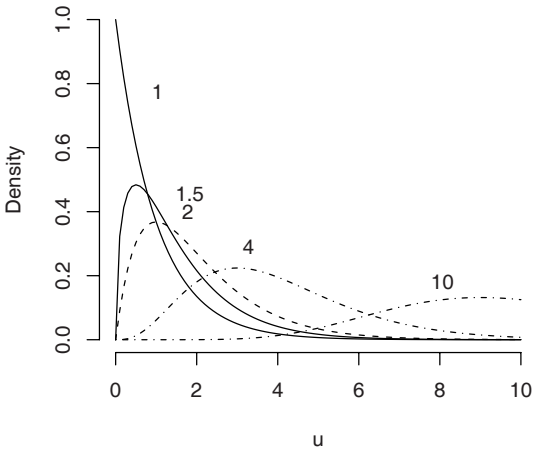


Fig. 8.6 Γ densities distribution for various shape parameters λ (shown by curve)

It can be challenging to use the Γ-distribution because the corresponding likelihood function does not have a closed form; i.e. we cannot write a function that calculates the likelihood value. Therefore, this distributional assumption can only be used by approximating the value of the likelihood function via numerical methods, and this turns out to be difficult and sometimes impossible.

What is the difference?

To compare the difference between the distribution of $e = v - u$ using half-normal distribution and the Γ-distribution for u, we can consider the densities for e as shown in Fig. 8.7, which corresponds to the plot in Fig. 7.5. However, here we only plot the density for e and not that for v and u, such that we can draw them all in one figure.

The overall picture of the figure on the left and right is the same, but if we look carefully, we can see some differences. The differences are not of empirical relevance, however; therefore it is advisable to use the simple model and the half–normal distribution.

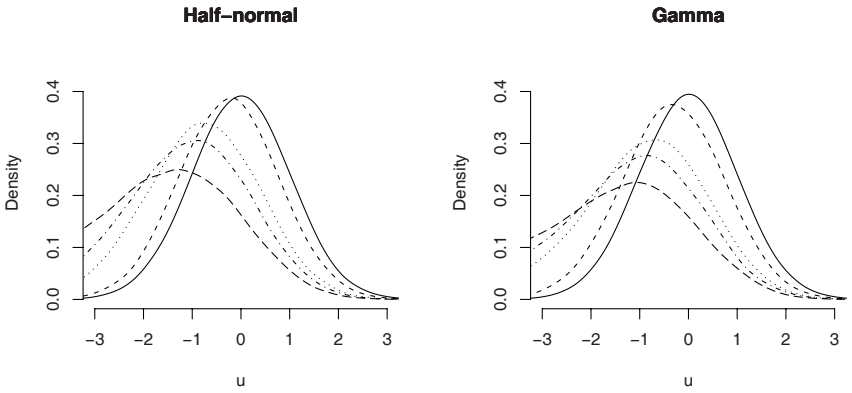


Fig. 8.7 Densities for $e = v - u$ for various distributions of u; the variances of u are the same in each corresponding curve on the left- and right-hand side

8.8 Biased estimates

One of the assumptions in ordinary regressions that carries over to SFA models is that the error terms must be independent of the regressors; i.e. u and v must be independent of x. If this condition is not fulfilled, then the estimates might be biased. This is well-known for ordinary regression models; for SFA models, we just provide a graphical illustration of the problem if u and one of the inputs x_i are correlated. For instance, perhaps firms that extensively use use x_i are more efficient than firms that use x_i on a more limited scale. This situation is shown in Fig. 8.8. The

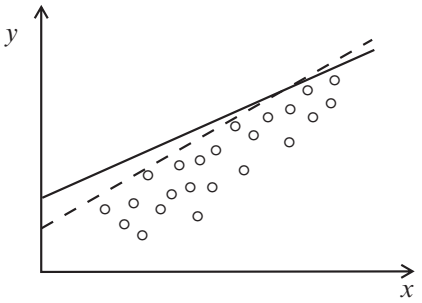


Fig. 8.8 Biased estimation when x_i and u are correlated

production function is shown as a solid line and the observations as small circles. For a small value of x, the circles are far below the solid line; these firms tend to very inefficient. Large x firms are close to but still below the solid line; these firms are very efficient. When we estimate the production line, we find the dashed line because in our estimation, we assume that x and efficiency u are independent and

therefore that the estimated line will follow the upper envelope of the circles except in instances of random noise. Therefore, the estimated slope (the dashed line) is different from the true slope (the solid line), and the estimated slope is biased.

This problem is a well known issue with models with random effects, and one solution is to use a fixed effect estimator. However, for cross-section data, the fixed effect estimator is of no use, and there is no solution to the problem except to remove the dependence from the us. A solution is possible if the data set is a panel data set.

For the data set on milk production, there does not seem to be a problem as seen from the scatterplot matrix in Fig. 8.9. The efficiency does not seem to depend on any of the regressors (i.e. vet, energy, and cows). The mean value-lines in the three middle plots in the lowest row is almost horizontal, showing that efficiency `te` does not depend on any of the three inputs, `vet`, `energy`, or `cows`. It does depend on the output `milk`, though, as seen in the bottom left plot; this was already discussed in connection with Fig. 7.15 on page 227.

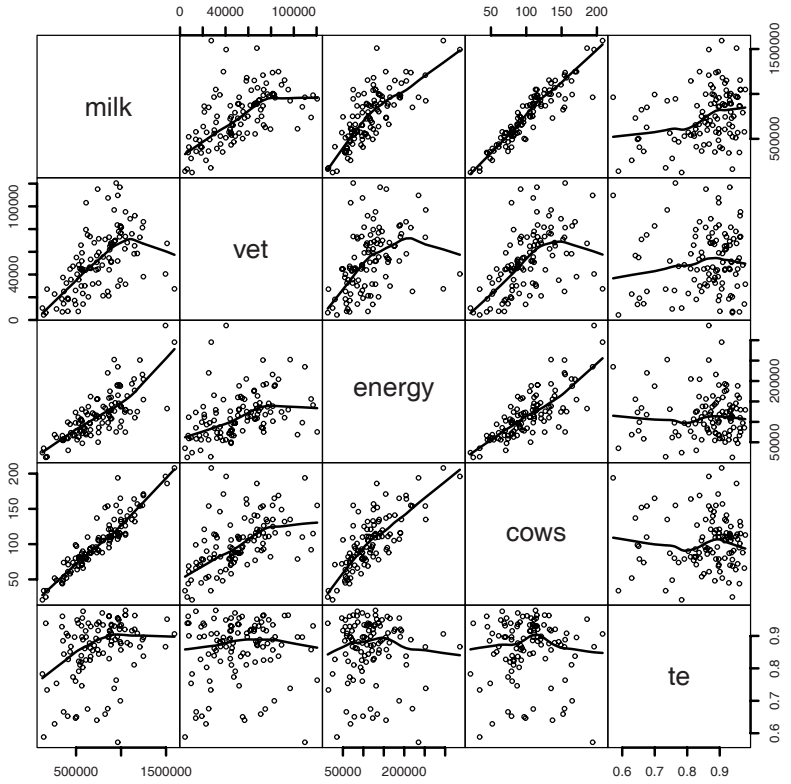


Fig. 8.9 Scatterplot matrix of efficiency, output, and input for production of milk

8.9 Summary

In this chapter, we have used distance functions descriptively and shown how we can estimate general distance functions using different classes of functional forms as part of an SFA approach. We have also included a brief review of general functional forms and of the translog function in particular.

We have shown that we can use a simple reinterpretation of the SFA production function model to also estimate cost functions.

As preparation for statistical hypothesis testing, we provided an informal overview of statistical inference based on likelihood theory, and we related this information to the numerical methods used to estimate the model. We showed how this can be used for ordinary *t*-tests in SFA models, more general linear hypothesis testing, and the most general form of test, the likelihood ratio test.

Lastly, we discussed the use of other distributions for the efficiency term, including the Gamma distribution, but we concluded that it was difficult to use these distributions in practical situations and recommended the continued use of the half–normal distribution instead.

We closed the chapter by mentioning some problems with the SFA models that are sometimes ignored in applications.

8.10 Bibliographic notes

The estimation of distance functions began after the appearance of translog functions and was first conducted in connection with SFA connection by Färe et al (1993). Since then, there have been many empirical applications of the method.

The translog function dates back to Christensen et al (1973) and has since been one of the most used parametric forms in empirical economics, both as a production function and as a cost function. Chambers (1988) has a chapter on the translog function and its merits in light of dual production theory.

Once we know the SFA production function, it is straightforward to also estimate the SFA cost functions.

The theory of statistical inference is the subject of many graduate courses in statistics and can be found in many statistical textbooks, including Silvey (1970), Rao (1973), Cox and Hinkley (1974), and Lehmann (1986).

The argument for using the half-normal distribution for efficiency over other distributions like the Γ-distribution because of its simplicity is found in Ritter and Simar (1997)

Panel data represent one way to handle some of the problems in SFA that we have mentioned. This idea is considered from a theoretical point of view in the book on SFA by Kumbhakar and Lovel (2000).

Duality in productions econonomic and Shephard's lemma are dicussed capably in Chambers (1988).

Chapter 9
Merger Analysis

9.1 Introduction

The quantitative literature on productivity has focused mainly on measuring the efficiency of individual firms and organizations. The results demonstrate how much can be gained by individual improvements, by learning best practices and by designing appropriate incentive schemes at the firm level.

This chapter expands our perspective and generalize the analytical techniques to the study of efficiency at the sector level. We will measure what can be gained by improving the structure of a group of firms and discuss mechanisms to accomplish this. We call this structural efficiency. It concerns the basic problem of coordination in a structure of multiple firms, i.e., the extent to which the right firms at the right locations are producing the right products at the right time. We concentrate first on horizontal mergers and then on reallocations of resources and services across firms and over time.

There are many reasons to be interested in structural efficiency. First, political decisions often affect—directly or indirectly—structural aspects of a sector. Agricultural policy, for example, has always affected structural development—and often the structural implications have been important considerations in designing policy. Second, the losses from suboptimal structures may be substantial. Analyses of several sectors, from fishery to hospitals, show that inadequate allocations and structures may be just as costly as individual inabilities to reach best practices. Third, instruments designed to improve individual performance may have negative structural impacts. In the energy sector, network companies with natural monopoly positions are routinely regulated by so-called revenue cap schemes that calculate the allowed revenue for a given company. This is a prime example of the use of state-of-the-art techniques like DEA and SFA in high stakes environments, and we will discuss it in details in Chap. 10, but it is also an example of possible negative structural impact. Unless the regulator is careful, the regulation can have adverse effects on structural development by making it unattractive for firms to cooperate or merge.

P. Bogetoft and L. Otto, *Benchmarking with DEA, SFA, and R*, International Series in Operations Research & Management Science 157, DOI 10.1007/978-1-4419-7961-2_9,

For a more specific motivating example, consider a competition authority responsible for accepting or rejecting a merger between two major companies. The decision requires careful examination of the pros and cons of the merger from a social point of view. The merger may limit competition and increase consumer prices. Oligopoly models are often used to estimate such market aspects. On the other hand, a merger may also lead to synergy effects that can reduce costs and improve quality. The methods in this chapter can be used to quantify these potential gains. In fact, some of the measures covered below are already in use by regulators and competition authorities in, for example, in Norway and the Netherlands.

9.2 Horizontal mergers

In this section, we develop models of the overall potential gains from the horizontal integration of two similar firms. We then decompose the gains and discuss a few refinements of the models.

Let us start by briefly recalling the idea of individual performance evaluation.

As we have seen again and again in this book, a simple way to think of an organization is as a transformation of multiple inputs (x) into multiple outputs (y). In the case of a hospital, for example, we may think of the inputs as numbers of doctors and nurses, while the outputs could be the number of treatments provided and the capacity provided as a buffer against uncertainty.

To evaluate the performance of a specific organization P^1, we need some benchmark against which to evaluate its production (x^1, y^1). In general terms, we may think of this benchmark as the technology T, alternatively described via the input sets $L(y)$ or the output sets $P(x)$. To evaluate the performance of organization P^1, we compare its resource usage and service provision against the technology T. If it is possible to produce more outputs than y^1 using fewer inputs than x^1, we say that P^1 is inefficient, and we may, for example, measure the amount of inefficiency by the Farrell measure of the input efficiency E^1 or the output efficiency F^1. The larger the distance from the frontier of the technology, the smaller the value of E^1 or the larger the value F^1 and the more inefficient the organization P^1 is.

We can use the same logic in evaluating merged entities as in evaluating individual entities. The larger the distance to the frontier, the more inefficient the merged firm is. Being inefficient represents a loss. On the other hand, being inefficient also suggests possibilities for improvement. This leads to the *basic idea* of this chapter. *Corporate synergy* occurs when corporations, through their interactions, are able to produce more services with a given set of resources, or to produce a given set of services with less resources. We can therefore capture the synergies from a merger by the increase in improvement potential when we move from independent to joint operations.

9.2.1 Integration gains

Figure 9.1 illustrates a classical horizontal merger. Two firms (or production entities) P^1 and P^2 individually transform vectors of inputs (resources), x^1 and x^2, respectively, into vectors of outputs (services), y^1 and y^2. Observe that we do not assume that they use exactly the same input and output types because we can always allow the values of some of the dimensions of the vectors x and y to be 0.

If the two firms integrate but continue to operate as two independent entities, they transform the vector of inputs $x^1 + x^2$ into the vector of outputs $y^1 + y^2$. To evaluate the potential efficiency gains from the merger, we can therefore evaluate the efficiency of the latter transformation, i.e., the use of $x^1 + x^2$ to produce $y^1 + y^2$.

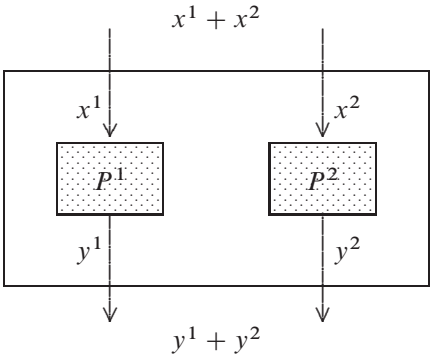


Fig. 9.1 Horizontal integration

Of course, the integration of more than two firms is also possible. Hence, in general, we consider a situation where a collection of firms indexed by some set H are integrated. The full set of firms about which we have information is denoted $K = \{1, \ldots, K\}$. In general, the set H is a subset of K, but this need not be the case. Hence, K could be all Danish hospitals and H could be all hospitals in the capital city of Copenhagen, or H could be all Swedish hospitals in Malmö, close to Copenhagen. Note that we use K and H to denote both numbers of firms and sets of firms in this chapter. This is a convenient notation, and the specific meanings of K and H will always be clear.

The merged firm is denoted P^H. Direct pooling of the inputs and outputs gives a firm that uses $\sum_{k \in H} x^k$ to produce $\sum_{k \in H} y^k$. This corresponds to having a completely decentralized (or compartmentalized) organization where the decentralized firms correspond to the H-firms. The inefficiency of the directly pooled production plan

$$\sum_{k \in H} x^k = \text{combined inputs}$$
$$\sum_{k \in H} y^k = \text{combined outputs}$$

is a measure of the improvement potential in the merged firm and can therefore be interpreted as the overall potential gains from the merger.

To illustrate this reasoning, consider Fig. 9.2. Two firms A and B have been technically efficient in the past, as indicated by the fact that they are located on the efficient frontier, the production function, ex ante. If they integrate but do not utilize the new synergies (in the illustration, the economies of scale), they would spend $(x^1 + x^2)$ to produce $(y^1 + y^2)$, as indicated by the point A+B. This is, however, a technically inefficient combined production because there are feasible productions to the northwest of $A + B$, i.e. it is possible to find alternative productions that use fewer inputs to produce more outputs, as reflected by the Potential Improvement *PI* set.

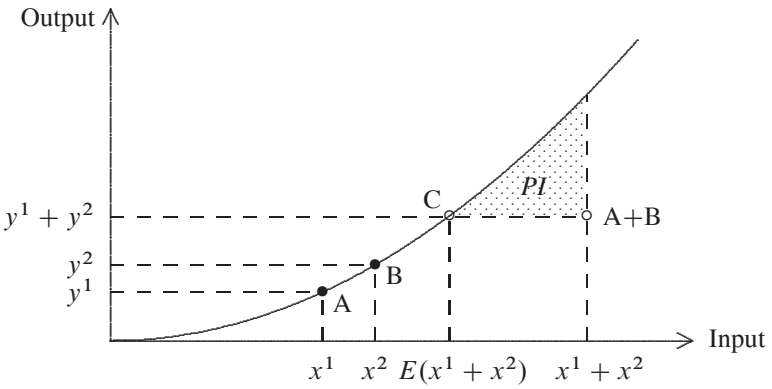


Fig. 9.2 Overall gains from horizontal integration

The possibilities for improvement can be summarized in different ways.

The simplest way is to use the Farrell measure on the input side. The Farrell input measure reduces to a simple comparison of the horizontal lengths of $A + B$ and C, and we see that the aggregate input consumption can be scaled down by the factor E.

Instead of focusing on the input (cost reductions), we could use the Farrell measure on the output side. This would stress the possibilities to increase outputs with a factor F. A more complete picture of the savings potential is to consider all point northwest of $A + B$, i.e. the set *PI*. Any such point can be generated, for example, by a directional distance function approach, by varying the improvement direction. In this case again, we obtain a score to measure the possible gains, namely, the excess e. In applications, the exact measure may be less important than the ability to investigate the set of potential improvements. This is the approach taken, for example,

in the interactive benchmarking approach used by Danish waterworks and Dutch hospitals, cf. Sect. 2.5.

Formally, a radial Farrell like input based measure of the *potential overall gains from merging* of the H-firms is

$$E^H = \min\{\, E \in \mathbb{R}_+ \mid \big(E \sum_{k\in H} x^k, \sum_{k\in H} y^k\big) \in T \,\}$$

such that E^H is the maximal proportional reduction in the aggregated inputs $\sum_{k\in H} x^k$ that allows the production of the aggregated output profile $\sum_{k\in H} y^k$.

If $E^H < 1$, the merger produces savings, and if $E^H > 1$, the merger is costly. A score of $E^H = 0.8$ would suggest that 20% of all inputs could be saved by integrating the firms in H. Likewise, a score of $E^H = 1.3$ would suggest that an integration would necessitate 30% more of all the resources.

So far, we have made no assumptions about the technology T. In practice, T must be estimated, and we can use any of a number of methods to do so, including the DEA and SFA approaches stressed in this book. We provide some illustrations of this below, but it is important to emphasize that the conceptual ideas do not rely on any specific estimation method.

In some situations, the above problem may have no solution at all. However, it is always feasible if T satisfies additivity. In particular, it is therefore feasible in the DEA models: CRS, IRS, and FDH.

One could of course measure the potential gains using many other indices. In particular, one could perform all of the evaluations and decompositions below on the output side.

Also, one can look at the potential improvements as a set, PI given by

$$PI^H = \big\{\, (x, y) \in T \mid x \le \sum_{k\in H} x^k \text{ and } y \ge \sum_{k\in H} y^k \,\big\},$$

i.e. PI^H is the set of feasible productions that use no more inputs to produce no less outputs, as illustrated in Fig. 9.2. One could also use a directional distance function approach by solving a problem like

$$e_d^H = \max\big\{\, e \in \mathbb{R}_+ \mid \big(\sum_{k\in H} x^k, \sum_{k\in H} y^k\big) + e(-d_x, d_y) \in T \,\big\}.$$

Here $d_x \in \mathbb{R}_+^m$ is the direction in input space, and $d_y \in \mathbb{R}_+^n$ is the direction in output space that we want to reduce and expand. The directional distance e is a measure of the number of times we can introduce the improvement packages $(-d_x, d_y)$.

In the following, we restrict ourselves to the Farrell approach. We note, however, that the decompositions developed below can be given a parallel treatment with directional distance functions—the main difference is that we get additive instead of multiplicative decompositions.

The overall potential gain from a merger, i.e., E^H, is an interesting starting point. It represents a best case, an upper limit scenario, that can be used, for example,

by competition authority as a first test to see if the efficiency gains can possibly outweigh the competitive effects.

However, the overall measure is optimistic and crude and requires refinements in several directions.

First of all, some of the gains could possibly be obtained without mergers, and can therefore not be associated directly with the mergers. We decompose the gains into learning, scope and scale effects to account for this.

Second, the overall gains may be too optimistic because there may be restrictions on the controllability and transferability of the resources and services. We discuss this below as well.

Third, one can question the assumption that the merged entity will be technically efficient, given that firms even in highly competitive industries show inefficiencies. We show also how to relax this assumption.

9.2.2 Disintegration gains

We have so far looked at the likely overall impact of merging two or more firms. The existence of potential positive synergies depends on the details of the firms being merges and the details of the underlying technology. This means that it is sometimes less resource consuming to operate two independent firms rather than one joint firm. This is not surprising because the coordination and motivation tasks in large organizations may be considerable. This also explains why we sometimes see different divisions of a joint enterprise operate independently, e.g., as individual profit-centers.

In fact, we can use the same logic as above to investigate the potential gains from the disintegration of large entities. For illustration, assume that we consider splitting up a firm (x^H, y^H) into two firms (x^1, y^1) and (x^2, y^2). Now, if it is possible to find feasible plans for the individual firms, i.e.

$$(x^1, y^1) \in T$$
$$(x^2, y^2) \in T$$

such that the individual firms together use less resources to produce more services

$$x^H \geq x^1 + x^2$$
$$y^H \leq y^1 + y^2,$$

then we can look at these reduced inputs and expanded outputs as an indication of the potential gains from disintegration. That is, we can measure the *potential gains from disintegration* as

$$
\begin{aligned}
E = \min_{(x^1,y^1),(x^2,y^2)} \; & E \\
\text{s.t.} \quad & E x^H \geq x^1 + x^2 \\
& y^H \leq y^1 + y^2 \\
& (x^1, y^1) \in T \\
& (x^2, y^2) \in T.
\end{aligned}
$$

In other words, we seek to find two feasible production plans (x^1, y^1) and (x^2, y^2) that together are able to produce at least the same output as (x^H, y^H), and to ensure the largest possible proportional reduction of all inputs.

Note that if $E < 1$, there is a potential saving involved. This would typically happen when (x^H, y^H) operates somewhat above the optimal scale size. It is of course also possible that $E > 1$, suggesting a net cost of forcing a disintegration. Such analyses can therefore be used to make trade-offs between required disintegrations to increase competition and losses in the economic efficiency of production.

9.3 Learning, harmony and size effects

Our measures of the potential overall merger gain from a merger encompass several effects. We now decompose them into technical efficiency, scale, and mix effects and discuss the organizational relevance of this decomposition. We first illustrate the ideas before presenting more formal derivations of the effects.

We can identify at least three sources of improvement.

One is *technical efficiency or learning* and is associated with the ability to adjust to best practices. Consider a horizontal merger of A and B illustrated in Fig. 9.3 below.

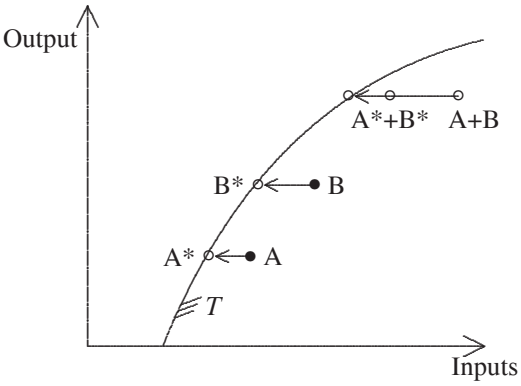


Fig. 9.3 Learning or technical efficiency effect

If the organizations merge but operate as they have done in the past, we see that there are considerable saving potentials, as represented by the distance of $A + B$ to the production possibility set. One can argue, however, that a considerable share of these potential gains were also available on an individual basis if the individual entities had optimized their businesses as represented by the dots A^* and B^*. If businesses A^* and B^* integrate, this would lead to the aggregate dot $A^* + B^*$, where the potential savings are considerably less than in $A + B$. We refer to this as a learning or technical efficiency effect and say that this is not—at least not completely—associated with the merger.

Another source of potential savings, called the *scope or harmony effect*, is associated with the mix of resources used and the mix of services provided. To illustrate this, consider two firms, e.g., hospitals, with the same levels of output and input requirements corresponding to the $L(y)$ curve as illustrated in Fig. 9.4 below.

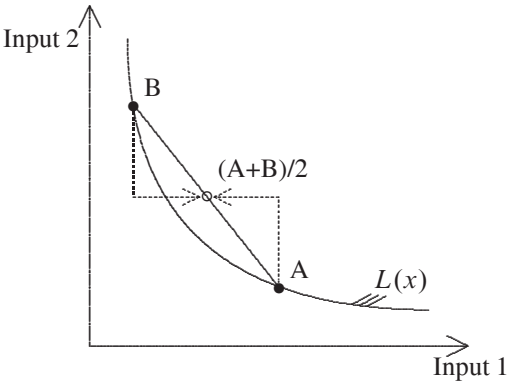


Fig. 9.4 Harmony or scope effect

We see that A is quite Input 1 intensive while B is Input 2 intensive. It is clear, however, that neither of the factor mixes may be optimal—at least, they cannot be optimal simultaneously. We see that the rate of substitution between Input 1 and Input 2 is different in the two firms. In A, a large amount of Input 1 is required to compensate for the loss of extra Input 2, while in B, many Input 2 units are required to compensate the loss of one Input 1. This means that there are possibilities to improve by moving some Input 2 from B to A and some Input 1 from A to B. If we move the factors as indicated with the dashed lines, both firms end up at $(A + B)/2$. We see that there are now possibilities for each of the firms to save. Of course, similar possibilities exist on the output side, i.e. by moving some obligations from A to B and other obligations from B to A, we can get service combinations that requires less resources to produce or that match the existing factor combinations in a better way. We talk about this effect as the *harmony or the scope effect*. Again, the point is that if independent enterprises just cooperated somewhat, they could gain and improve their premerger performance, making the pure gains from a merger less.

In addition to these effects, a merger will also have an impact on the scale of operation. This leads to the so-called *scale or size effect*. We have already illustrated this in Fig. 9.2. Note that in the illustration, both firms are fully efficient individually, such that there can be no learning effect. Likewise, we have only one input and one output such that there can be no harmony effect. In the case of a technology with economies of scale, it is attractive for firms to be large since this allows them to produce at lower average costs. Of course, the scale effect need not be positive—it depends on the underlying technology whether the increase in scale is favorable or not.

The three effects above, the learning, harmony and size effects, determine the combined effect of a merger. We will now formalize these concepts and then return to their interpretations and integration.

To adjust the overall merger gains for the *learning effect*, we can project the original firms to the production possibility frontier and use the projected plans as the basis for evaluating the remaining gains from the merger. Thus, we project (x^k, y^k) onto $(E^k x^k, y^k)$ for all $k \in H$, where E^k is the standard efficiency score for the single kth Firm, and use the projected plans $(E^k x^k, y^k)$, $k \in H$, as the basis for calculating the *adjusted overall gains* E^{*H} from the merger:

$$E^{*H} = \min\left\{ E \in \mathbb{R}_+ \mid \left(E \sum_{k\in H} E^k x^k, \sum_{k\in H} y^k\right) \in T \right\}.$$

If we set

$$LE^H = \frac{E^H}{E^{*H}},$$

we get $E^H = LE^H \cdot E^{*H}$, where $LE^H \in [0, 1]$ indicates what can be saved by individual technical efficiency adjustments in the different firms in H.

Assuming that individual technical inefficiencies have been dealt, we are left with the scaling or size effect, on the one hand, and the harmony, scope or mixture effect, on the other hand.

To formalize the *harmony gains* HA^H we examine how much we can reduce the average input in the production of the average output:

$$HA^H = \min\left\{ HA \in \mathbb{R}_+ \mid \left(HA \frac{1}{H} \sum_{k\in H} E^k x^k, \frac{1}{H} \sum_{k\in H} y^k\right) \in T \right\},$$

where H is both the set and the number of elements in the set of firms in the merger. We look at the average input and average output because we do not yet want the expansion of size to come into play. Using the average is most relevant if the firms in H are not too different in size to begin with. If the sizes differ considerably, we may be picking up scale effects, for example, if some firms are larger than and some are smaller than the most productive scale size. Note that $HA^H < 1$ indicates a potential savings due to improved harmony, while $HA^H > 1$ indicates a cost of harmonizing the inputs and outputs.

Next, we capture the *size gains* SI^H by asking how much could have been saved by operating at full scale rather than average scale:

$$SI^H = \min\left\{ SI \in \mathbb{R}_+ \mid \left(SI \cdot HA^H \sum_{k\in H} E^k x^k, \sum_{k\in H} y^k\right) \in T \right\}.$$

The rescaling is advantageous, $SI^H < 1$, if we have economies of scale, and costly, $SI^H > 1$, if the returns to scale property does not favor larger firms.

Using the above notions of learning *LE*, harmony *HA* and size *SI* effects, we get our *basic decomposition*

$$E^H = LE^H \cdot HA^H \cdot SI^H.$$

The learning or technical efficiency measure LE^H captures what can be gained by making the individual firms efficient. The remaining potential savings, E^{*H}, are created by the harmony or scope effect HA^H, and the size or scale effect SI^H.

How to calculate the measures exactly in general multi-input multi-output technologies depends on how they are represented. We will return to this in the next section, where we discuss both parametric and nonparametric implementations and where we also provide R code to support the calculation.

We note that the learning effect is always weakly positive (nonnegative) in the sense that $LE \leq 1$, such that there are potential savings $1 - LE$.

To know the signs of the harmony and size effects a priori, we need more assumptions about the underlying technology. For example, if the technology is convex, the harmony effect is always weakly positive $HA \leq 1$, while the size effect may or may not favor a merger in a convex technology. In a convex technology that also satisfies the assumption of constant or increasing returns to scale, the size effect is always positive.

9.3.1 Organizational restructuring

The decomposition of the potential gains from merging firms into a technical efficiency measure, a harmony measure and a size measure is important because full scale mergers are typically not the only available organizational option, and alternative organizational changes may be easier to implement. In particular, we suggest the following guidelines for organizational restructuring:

Low learning measure *LE*

One could let the inefficient firms learn from the practices and procedures of the more efficient ones. If the problem is not a lack of skills, but rather a lack of motivation, one could improve the incentives, e.g., by using relative performance evalua-

tion and yardstick competition based on the technical efficiency measures, cf. Chap. 10. Of course, if the problem is the scarcity of managerial talent, it may still be necessary to make a genuine merger to transfer control to the more efficient administrative teams and thereby improve the managerial efficiency (X-efficiency). Another effect of a genuine merger emphasized by practitioners is the fact that a merger is a change event where established rules and procedures are being re-evaluated and improved. This is because every organization has some slack, which it is difficult to reduce under normal conditions.

Low harmony measure *HA*

One could consider reallocating the inputs and outputs among the firms to create more powerful input mixes and more easily produced output mixes. This can be done (a) inside a hierarchy, (b) by long term contracts or, perhaps, (c) by creating a market for key inputs and outputs.

Low size measure *SI*

In this case, full scale mergers may be the only alternative. If we need large amounts of fixed capital, highly specialized staff, long run-lengths or simply a critical mass to obtain sufficient returns from scale, it may be relevant to merge. In addition, and perhaps most important, this may be relevant if reallocation through contracts or a market is associated with too many transaction costs to make it attractive, cf., the general discussion of optimal firm size in the industrial organization literature.

9.3.2 Rationale of the harmony measure

The decomposition developed above gives a natural way to define and distinguish between the technical efficiency, the size and the harmony effects.

On the other hand, one must acknowledge that decompositions are in general ambiguous in the sense that one can decompose them in different orders and get different measures. It is therefore always important to look for more profound rationales.

An important such rationale for the harmony measure is that *with a convex, free, disposable technology, the harmony effect measure is the most that can be gained by any kind of reallocation between the firms in* H. Assume that we were to pick new inputs and outputs (x^{*k}, y^{*k}) for each $k \in H$ such that the total inputs and outputs stay feasible, $\sum_{k\in H} x^{*k} \leq \sum_{k\in H} x^k$ and $\sum_{k\in H} y^{*k} \geq \sum_{k\in H} y^k$, and such that all of the new productions are feasible, $(x^{*k}, y^{*k}) \in T$. Now, the largest possible savings, of $\sum_{k\in H} E^k x^k$, is precisely the harmony effect. That is, assuming the free disposability and convexity of the technology T, *HA* is also the solution to

$$\begin{aligned} \min_{((x^{*k},y^{*k}))_{k\in H}} \quad & h \\ \text{s.t.} \quad & h\sum_{k\in H} E^k x^k \geq \sum_{k\in H} x^{*k} \\ & \sum_{k\in H} y^k \leq \sum_{k\in H} y^{*k} \\ & (x^{*k}, y^{*k}) \in T. \end{aligned}$$

9.3.3 Decomposition with a cost function

Before concluding this discussion of the basic decomposition, it may be useful also to illustrate it in the single input (cost) multiple output context. Thus, let

$$c(y) = \min\{x \in \mathbb{R}_+ \mid (x, y) \in T\}$$

be the underlying cost function, which gives an alternative representation of the underlying technology. We then have

$$\begin{aligned} E^H &= c\Big(\sum_{k\in H} y^k\Big) \Big/ \sum_{k\in H} x^k \\ E^{*H} &= c\Big(\sum_{k\in H} y^k\Big) \Big/ \sum_{k\in H} c(y^k) \\ LE^H &= \sum_{k\in H} c(y^k) \Big/ \sum_{k\in H} x^k \\ HA^H &= c\Big(\frac{1}{H}\sum_{k\in H} y^k\Big) \Big/ \Big(\frac{1}{H}\sum_{k\in H} c(y^k)\Big) \\ SI^H &= c\Big(\sum_{k\in H} y^k\Big) \Big/ \Big(H\, c\Big(\frac{1}{H}\sum_{k\in H} y^k\Big)\Big). \end{aligned}$$

As these expressions show, the learning effect LE^H measures the reduction in costs if everyone learns best practices but remains an independent entity, the harmony effect HA^H measures the minimal cost of the average output vector compared to the average of the costs corrected for individual learning, and the size effect SI^H measures the cost of operating at the full (integrated) scale compared to the average scale of the original entities.

9.4 Implementations in DEA and SFA

Technically, to calculate the overall and decomposed measures of potential gains from mergers, we simply need the ability to calculate the efficiency of different input–output combinations (x^*, y^*), namely, to do the following:

Overall potentials E^H*:* Evaluate the direct combination of the firms involved in the merger, i.e. find the efficiency of $\left(\sum_{k\in H} x^k, \sum_{k\in H} y^k\right)$

Individual potentials E^k*:* Evaluate individual learning potentials, i.e. find the efficiency of $(x^k, y^k), k \in K$

Pure merger potentials E^{*H}*:* Evaluate the merged firm after individual learning, i.e. find the efficiency of $(\sum_{k\in H} E^k x^k, \sum_{k\in H} y^k)$

Learning effect LE^H*:* Compare the overall potential to the pure merger potential, i.e. calculate $LE^H = E^H / E^{*H}$

Harmony effect HA: Evaluate the average efficient firm, i.e., find the efficiency of $\left(\frac{1}{H}\sum_{k\in H} E^k x^k, \frac{1}{H}\sum_{k\in H} y^k\right)$

Size effect SI: Evaluate the efficiency of the upscaled average efficient firm, i.e., find the efficiency of $\left(HA^H \sum_{k\in H} E^k x^k, \sum_{k\in H} y^k\right)$, or simply find the size as a residual, $SI^H = E^{*H}/HA^H$.

Therefore, we can use any benchmarking approach that allows us to calculate the efficiency of firms against some fixed technology. Note that the underlying technology does not change in these calculations. The fact that we look at a combined firm like $\left(\sum_{k\in H} x^k, \sum_{k\in H} y^k\right)$ does not mean that we assume that it is feasible or inside the technology. It may be outside, suggesting that it would be costly to undertake the merger.

This also means that by repeated use of any software able to calculate super-efficiencies, we can do the calculations. We need super-efficiencies since the firms we evaluate should not affect the technology.

A simple way to implement these calculations is to use the *R* programming environment. The principal steps are given in Table 9.1.

A few notes may help interpret these code lines.

The inputs and outputs of the firms whose mergers we want to analyze are given in the matrices X and Y. The merge matrix M has one row for each of the mergers we want to analyze, and the columns represent the initial firms. In a row, a 1 indicates that the corresponding column firm is included in the merger, while a 0 indicates that it is excluded.

The efficiency of any input output combination (x^*, y^*) is calculated with the function $\text{Eff}(x^*, y^*)$. This function can be defined as the solution to a linear programming problem, as is done in DEA, or as a line search problem depending on the underlying model of the technology. In the case of a single input technology

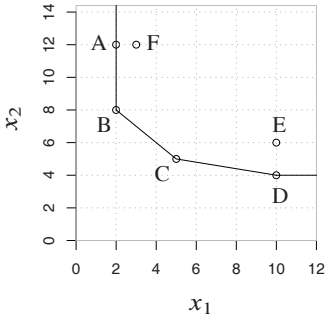
Table 9.1 Schema for efficiency analysis of mergers in R

```
# Input data, K times m matrix
X <- Input data with firms in rows
# Output data, K times n matrix
Y <- Output data with firms in columns

# Merge matrix with K collums.
# Each row defines a merger, =1 to include the firm.
M <- aggregation matrix with information on which mergers to analyze

# Efficiency measure
# Mergers should be measured against a fixed technology set,
# defined via a DEA model or a parametric model
Eff(X*,Y*) <- function to calculate efficiency of (X*,Y*)

# Input and output of merged firms
Xmerger <- M \%*\% X
Ymerger <- M \%*\% Y

# Potential overall gains, efficiency after merger
E <- Eff(Xmerger, Ymerger)

# Individual efficiencies before merger
e <- Eff(X, Y)

# Inputs of individual firms projected on efficient frontier
Xeff <- diag(e) \%*\% X

# Inputs of merged firms after elimination of individual inefficiency
Xmerger_proj <- M \%*\% Xeff

# Pure gains from mergers
E_star <- Eff(Xmerger_proj, Ymerger)

# Learning effect
LE <- E/E_star

# Inputs and outputs for merged firms in harmony calculation
Xharm <- diag(1/rowSums(M)) \%*\% Xmerger_proj
Yharm <- diag(1/rowSums(M)) \%*\% Ymerger_proj

# Harmony effect, compoared to original technology set
HA <- Eff(Xharm, Yharm)

# Size effect
SI <- E_star/HA
```

represented by the cost function $C(y)$, it could simply be $C(y^*)/x^*$. We will give examples of how to define Eff(x^*, y^*) in the numerical example below.

Table 9.2 Data for numerical merger example

Firm	x_1	x_2	y
A	2	12	1
B	2	8	1
C	5	5	1
D	10	4	1
E	10	6	1
F	3	12	1



Note also that `Xmerger_proj <- M %*% Xeff` are the inputs of the merged firms after having eliminated individual inefficiencies. Here, `%*%` represents matrix multiplication.

The same inputs, scaled down by dividing with the number of firms in the mergers, are given in `Xharm <-diag(1/rowSums(M)) %*%X merger_proj`, and similarly for the outputs, except that, here, we do not make projections.

As the results of these calculations, we have in the E, LE, HA and SI vectors the values of E^H, LE^H, HA^H, SI^H for all the mergers H corresponding to rows in the M matrix.

9.4.1 Numerical example in R

Consider the same problem as in Chap. 2. For convenience, the data and a graphical illustration are repeated in Table 9.2.

We now analyze three possible mergers, namely, $H = \{A, C\} = \{1, 3\}$, $H = \{E, F\} = \{5, 6\}$ and $H = \{A, C, F\} = \{1, 3, 6\}$. Assume that we measure efficiency using an IRS technology DEA model defined by the original six observations. Using the code schema above, we can now proceed as follows:

```
> library(Benchmark)
> xobs <- matrix(c(2, 2, 5, 10, 10, 3, 12, 8, 5, 4, 6,12), ncol=2)
> yobs <- matrix(rep(1,6), ncol=1)
> cbind(xobs, yobs)
     [,1] [,2] [,3]
[1,]    2   12    1
[2,]    2    8    1
[3,]    5    5    1
[4,]   10    4    1
[5,]   10    6    1
[6,]    3   12    1
# Individual efficiency before any merger
> dea(xobs,yobs,RTS="irs")
```

```
[1] 1.0000 1.0000 1.0000 1.0000 0.7500 0.6667
> # Mergers should be measured against the originale technology set,
> # therefore we must use xobs and yobs to determine the technology set.
> # In the definition of Eff below the use of XREF and YREF makes this happen.
> Eff <- function(X,Y){
+   e <- dea(X, Y, RTS="irs", ORIENTATION="in",
+            XREF=xobs, YREF=yobs)
+   return(e$eff)  # $
+ }
> grouping <- list(c(1,3), c(5,6), c(1,3,6))
> M <- make.merge(grouping,X=xobs)
> M
     [,1] [,2] [,3] [,4] [,5] [,6]
[1,]    1    0    1    0    0    0
[2,]    0    0    0    0    1    1
[3,]    1    0    1    0    0    1
> Xmerger <- M %*% xobs
> Ymerger <- M %*% yobs
> cbind(Xmerger,Ymerger)
     [,1] [,2] [,3]
[1,]    7   17    2
[2,]   13   18    2
[3,]   10   29    3
> E <-Eff(Xmerger,Ymerger)
> E
[1] 0.8333333 0.6451613 0.7692308
> e <- Eff(xobs,yobs)
> e
[1] 1.0000 1.0000 1.0000 1.0000 0.7500 0.6667
> Xeff <- diag(e) %*% xobs
> Xeff
     [,1] [,2]
[1,]  2.0 12.0
[2,]  2.0  8.0
[3,]  5.0  5.0
[4,] 10.0  4.0
[5,]  7.5  4.5
[6,]  2.0  8.0
> XmergerProj <- M %*% Xeff
> XmergerProj
     [,1] [,2]
[1,]  7.0 17.0
[2,]  9.5 12.5
[3,]  9.0 25.0
> Estar <- Eff(XmergerProj,Ymerger)
> Estar
[1] 0.8333333 0.9090909 0.8823529
> LE <- E/Estar
> LE
[1] 1.0000000 0.7096774 0.8717949
> Xharm <-diag(1/rowSums(M)) %*% XmergerProj
> Yharm <-diag(1/rowSums(M)) %*% Ymerger
> cbind(Xharm, Yharm)
     [,1]     [,2] [,3]
```

```
[1,] 3.50 8.500000    1
[2,] 4.75 6.250000    1
[3,] 3.00 8.333333    1
> HA <- Eff(Xharm,Yharm)
> HA
[1] 0.8333333 0.9090909 0.8823529
> SI <- Estar/HA
> SI
[1] 1 1 1
```

The above calculations can be simplified by using the function `dea.merge` from the package Benchmark, which does all the above calculations in just one function call:

```
> dea(xobs,yobs,RTS="irs")
[1] 1.0000 1.0000 1.0000 1.0000 0.7500 0.6667
> em <- dea.merge(xobs,yobs,M,, RTS="irs")
> em
$Eff
[1] 0.8333333 0.6451613 0.7692308

$Estar
[1] 0.8333333 0.9090909 0.8823529

$learning
[1] 1.0000000 0.7096774 0.8717949

$harmony
[1] 0.8333333 0.9090909 0.8823529

$size
[1] 1 1 1
```

We see that the overall efficiencies E (`em$Eff`) of the three mergers$\{A, C\}, \{E, F\}$ and $\{A, C, F\}$ are 0.83, 0.65, and 0.77, indicating potential gains from the mergers of 0.17, 0.35 and 0.33, respectively.

Now, for the two last mergers, part of these gains are learning effects because E and F are both inefficient on an individual basis. If we eliminate their individual efficiency before the merger, the pure merger efficiencies E^* (`em$Estar`) are 0.83, 0.91, and 0.88, with corresponding potential gains of 0.17, 0.09, and 0.12. Hence, what appeared to be the least promising merger, $\{A, C\}$, is now the most promising. The reason is that the learning inefficiencies have been eliminated in the other two mergers, and these inefficiencies are nontrivial because LE (`em$learning`) takes values of 1.00, 0.71, and 0.87, with corresponding individual learning potentials of 0, 0.29 and 0.13.

Now, the decomposition of the pure merger efficiency into the harmony and size effects is rather trivial in this example. Looking at the illustration in Table 9.2, it is clear that the efficient frontier will look like a CRS technology above the iso-quant (and that the estimated output for input combinations below the iso-quant will all be 0). Therefore, all mergers are operating at the CRS part of the frontier, and therefore,

the size efficiencies will all be 1 and the pure gains will all come from the harmony effect.

9.4.2 Mergers in a parametric model

If we used a different approximation of the technology, we would of course get different measures of the potential gains from the mergers. To illustrate this, let us assume that the underlying production possibilities are derived from a log-linear (Cobb-Douglas) function

$$y = 0.35\, x_1^{0.4}\, x_2^{0.4}.$$

The fully efficient output levels for the six firms in the example can therefore be calculated as 1.25, 1.06, 1.27, 1.53, 1.80, and 1.47. It follows that all firms are now initially inefficient, although B is close to efficient. Indeed, the input efficiency of a firm using x^* to produce y^* can be calculated as the solution E to

$$y^* = 0.35\,(E\, x_1^*)^{0.4}\,(E\, x_2^*)^{0.4} = 0.35\, E^{0.8} x_1^{*0.4}\, x_2^{*0.4}$$

i.e., we can get a closed form efficiency expression

$$E = \Big(\frac{y^*}{0.35\, x_1^{*0.4}\, x_2^{*0.4}}\Big)^{\frac{1}{0.8}}.$$

Now, the R calculations should look like this:

```
> library(Benchmark)
> X <- matrix(c(2, 2, 5, 10, 10, 3, 12, 8, 5, 4, 6,12), ncol=2)
> Y <- matrix(rep(1,6), ncol=1)
> # Define Eff
> Eff <- function(X,Y){
+    e <- (Y/(.35*apply(X^.4,1,prod)))^(1/.8)
+    return(array(e))
+ }
> grouping <- list(c(1,3), c(5,6), c(1,3,6))
> M <- make.merge(grouping,X=X)
> Xmerger <- M%*%X
> Ymerger <- M%*%Y
> E <-Eff(Xmerger,Ymerger)
> E
[1] 0.8098954 0.5775565 0.8612268
> e <- Eff(X,Y)
> e
[1] 0.7582446 0.9286562 0.7429249 0.5873337 0.4795560 0.6191041
> Xeff <- diag(e)%*%X
> Xmerger_proj <- M%*%Xeff
> Estar <- Eff(Xmerger_proj,Ymerger)
> Estar
[1] 1.079122 1.066941 1.224352
> LE=E/Estar
```

```
> LE
[1] 0.7505135 0.5413200 0.7034145
> Xharm <-diag(1/rowSums(M))%*%Xmerger_proj
> Yharm <-diag(1/rowSums(M))%*%Ymerger
> HA <- Eff(Xharm,Yharm)
> HA
[1] 0.9074296 0.8971869 0.9303062
> SI <- Estar/HA
> SI
[1] 1.189207 1.189207 1.316074
```

Note that as the technology is not a DEA technology the function `dea.merge` is not applicable here. In this case, however, one can simply follow the general schema from Tables 9.1 which is exactly what we did in the example.

We see that in this case, all mergers have learning potentials, but if we eliminate these, the pure merger gains will be negative: the E^{*H} values are in this case 1.08, 1.07, and 1.22, suggesting that it would require some 8% , 7% and 22% extra of both inputs to make the mergers $\{A, C\}$, $\{E, F\}$ and $\{A, C, F\}$, respectively. The decomposition shows that we would in all cases save resources by the harmony effect. The net cost of the mergers is due to the size effect. The production technology exhibits decreasing returns to scale. If we double the inputs, we only get $2^{0.8} = 1.74$ times the output, and this is of course a major disadvantage when we roughly double or triple the firm size.

Before closing this section, let us make two remarks.

9.4.3 Technical complication

First, in the general case of multiple inputs and outputs being evaluated in a parametric model, the calculation of the merger efficiencies may require some numerical estimation.

Without loss of generality, we may assume that we have estimated a Shephard input distance function $D_i(x, y)$ from the actual data. Using this, we can approximate the technology as

$$T^* = \{\, (x, y) \in \mathbb{R}_+^{m+n} \mid D_i(x, y) \geq 1 \,\},$$

and the calculation of the the efficiency of (x^*, y^*), i.e., $\mathrm{Eff}(x^*, y^*)$, therefore requires us to solve

$$\mathrm{Eff}(x^*, y^*) = \min\{\, E \mid D_i(Ex^*, y^*) \geq 1 \,\}.$$

With the usual Farrell measure E, this is particularly simple thanks to the relationship between the Farrell and Shephard distance measures, i.e., $E = 1/D_i$ in this case.

However, when we want to make other estimates, for instance, if we assume some inputs to be discretionary (variable) and others nondiscretionary (fixed), the

calculation of the efficiency is a little more complex because we must then solve

$$\min\{ E \mid D_i(E x_v^*, x_f^*, y^*) \geq 1 \},$$

and this equation may not have a closed form solution. Still, since D_i is monotone increasing (weakly increasing) in x_v, and therefore in the scalar E, this problem can be solved by a simple *line search*, e.g., by a *bisection* approach that does not require any further assumptions about the functional form of D_i. Similar methods are necessary if we want to measure and decompose gains from mergers using a direction distance function in the parametric case.

9.4.4 Methodological complication

Second, in the case of SFA, a further complication concerns the interpretation of the efficiencies. If we estimate a SFA cost or production function, or even a more general distance function, we know that actual observations will deviate from the estimate for two reasons. One is random noise and the other is inefficiency.

If we use the estimated function as above, we will basically assign all deviations from the cost, production or distance function to the inefficiency term. If we do so, we get the measures and the decompositions above.

However, one could argue that this is not the correct way because it conflicts with the underlying idea of SFA. Instead, to evaluate an observation (x^*, y^*), one could calculate the deviation ϵ and then find the conditional expectation of the inefficiency given this deviation. Although this would seem to be the correct approach, the interpretation is not as simple as in the case of a pure inefficiency model. The reason is that if we take a SFA model and estimate the efficiency of a given firm, and then we create a new firm that is the same as the original firm except without the estimated inefficiency, then this new firm would not necessarily be efficient in the sense that the conditional efficiency would now be 1. Hence, we cannot say that E^{*H} is the efficiency of the combined firm after having eliminated the individual inefficiencies.

Therefore, it is not yet clear exactly how best to proceed in the SFA case.

9.5 Practical application: Merger control in Dutch hospital industry

The evaluation of the potential gains from mergers and the decomposition of these gains into learning, harmony and size effects are used by the Dutch health authority NZa, among others, to decide whether to grant merger permissions and under which conditions. The a priori assessment of mergers and agreements requires careful consideration because the negative and positive effects of mergers stem from a number of different sources, and they have to be weighed against each other. An additional

complication lies in the fact that the calculations and estimates concern the future. We often have to compare two hypothetical situations: (1) what are the probable effects of the merger; (2) what would happen in the market without a merger?

The market impact has traditionally been estimated using models of imperfect competition, e.g., industrial organization models of Cournot competition. There exists a relatively well-developed set of approaches for this purpose.

In 2008, NZa initiated a project to improve the estimation of the possible efficiency gains. The aim was to develop methods to compute potential gains from horizontal and vertical mergers. The mergers of primary concern to NZa were between pairs of hospitals and between hospitals and insurance companies. In general, the recent liberalization of the Dutch healthcare sector has led to a number of mergers between healthcare and related institutions.

In this project, we therefore developed a series of new measures and software programs to implement the ideas on real data. In particular, we developed a version of interactive benchmarking that allows NZa easily to explore the full potential improvement set *PI* from a merger.

As part of the study, we investigated the potential gains from the horizontal integration of Dutch hospitals. Specifically, we used cost and production data from 97 hospitals in 2006 to estimate best practice DEA and SFA models. We used physical distance information to determine all potential pairs of merges of two hospitals with a maximal distance of 10 km. There are 37 possible such mergers. For each of these pairs, we then evaluated the total potential gain and its decomposition into learning efficiency, harmony and size potentials.

In hospital models, detailed output descriptions are defined using diagnosis related groupings, DRGs. Most countries work with systems that distinguish some 700-1000 different DRG outputs. In the Netherlands, they use a special variant that combines diagnosis and detailed treatment information to define close to 30,000 different products called DBCs. Each of these is assigned a price that is intended to reflect the cost of the DBC. Of course, from the cost accounting perspective as well as from the econometric point of view, one can naturally be skeptical as to the possibility of creating a meaningful cost break down at this level of detail.

In one analysis. we considered a grouping of these outputs into six turnover values or group volumes. That is, we depicted a hospital as transforming costs into six output categories. The categories are closely related to a Dutch proposal for how to define economically homogeneous specialty clusters, although a few adjustments were necessary to align with the available data. For each of the groups, we considered the total product of regulated DBCs. The value of this product, as evaluated with the DBC weights in use, was calculated to determine the group ”turnover” at regulated prices at a given hospital

$$y_j = \sum_{h \in j} p_h q_h,$$

where q_h is the number of DBC$_h$ produced, p_h is the regulated price of DBC$_h$, and where j is a group of DBCs. These turnovers are the outputs and cost drivers in our model of hospital service production.

The use of such weighted combinations of underlying heterogeneous productions is a common and useful way to reduce the number of degrees of freedom in any estimation approach. It basically implies that we accept the intra-group calibration, i.e. the relative prices inside the groups. Of course, one could continue like this and aggregate across the groups to obtain the total weighted output of a given hospital. This would mean that we also accept a priori that the inter-group calibration, i.e. the relative costs accross groups, implied by the DBC weights. Instead, we calibrate the relative importance of the cost driver groups using frontier models and the data available on total costs and total service productions from the 97 hospitals. A possible intermediate approach would be to add weight restrictions on the inter-group calibration, e.g., using assurance regions as discussed in Chap. 5. A summary of the data from the 97 hospitals is given in Table 9.3.

Table 9.3 Dutch hospital data in 1000 Euro

Statistics	Cost x	y_1	y_2	y_3	y_4	y_5	y_6
Average	124 891	66 986	15 230	4 375	16 454	1 539	131
Std.dev.	79 302	36 692	11 041	6 928	8 676	1 293	97
Min	23 598	8 115	1 771	2	1 804	129	2
Max	363 747	171 332	52 775	29 582	37 703	4 899	406

Using these data, we can estimate a linear average cost model as

$$c^k = \beta_o + \sum_{j=1}^{6} \beta_j y_j^k + \epsilon^k,$$

where c^k is the (relevant part of the) cost of hospital k, y_j^k is the production level for group j in hospital k and ϵ^k is a random noise term. A simple linear regression analysis gives an adjusted R-squared of 0.8656, i.e., the regression is able to explain a large share of the variation of costs by the six cost drivers in the model. The parameter estimates are given in Table 9.4. This simple regression suggests that it

Table 9.4 Average cost model

Statistics	β_0	β_1	β_2	β_3	β_4	β_5	β_6
Estimate	13 600 000	0.779	1.660	3.060	0.436	6.370	89.3
Std.dev	6 890 000	0.3	0.6	0.6	0.8	3.3	37.
t value	1.97	2.82	2.65	5.23	0.53	1.92	2.42
$\Pr(\beta_i > t)$	0.052	0.006	0.009	0.000	0.596	0.058	0.018

may be worthwhile to consider the calibration of the weights in the DBC system.

If the weight system is well calibrated we would expect the β values to almost the same. In particular, the regression analysis suggests that the weights in groups 2, 3, 5 and, in particular, 6 may be set somewhat below the real costs while the other groups have slightly boosted values. We acknowledge of course that the cost and product definitions and the data set used here are not sufficient to determine final conclusions as to the relevance of the actual DBC weights.

We next estimated a series of frontier models of the cost function, i.e., the costs as a function of the 6 outputs or cost drivers, $y_1, \ldots, y_6$:

$$c = C(y_1, \ldots, y_6).$$

The model specification, i.e., the inputs and outputs defined, was tested using both SFA and DEA approaches. In each class, we estimated a range of possible specifications to get an impression of the sensitivity of the results to the specification of the model. In the SFA framework, we estimated linear, log-linear, translog, normed linear and normed log-linear specifications of the mean structure and truncated normal distribution for the inefficiency error term. In the DEA framework, we made estimates using the scale assumptions CRS, DRS, IRS and VRS returns to scale. Specific runs were also made with a bias-corrected DEA model, including confidence interval, $[c1, c2]$, for the bias corrected efficiencies. A summary of the preliminary results on the Farrell input efficiencies E for the sample data is provided in Table 9.5. For each estimation methods, the table gives information about the mean Farrell efficiency, the standard deviation of the the Farrell efficiencies, the number of fully efficient hospitals and the lowest Farell efficiency among all hospitals.

As we can see from the summary of different estimations, the level of cost inefficiency $(1 - E)$ in the Dutch hospital sector is 10–20 percent in most specifications. The interpretation of this result is that if everyone learned best practices, the total costs could be reduced by 10–20% without changing the organization of the sector.

The scale inefficiency, dea-se, is approximately 7% in the DEA models, suggesting that some 7% could be saved if everyone adjusted to optimal scale size.

Although our aim in producing these test models was not to develop an authoritative cost model of Dutch hospitals, a few notes on these levels are useful. As a first quantification of cost inefficiency, compared with other sectors, the estimated cost inefficiency is not alarming. In fact, the results suggest considerable possibilities to save due to the large underlying costs in absolute euros, but in relative terms, one finds similar potential savings in many other sectors, both regulated and more competitive. It should on the other hand be observed that this level of estimated inefficiency may also reflect the way the DBCs are priced. Because they are intended to reflect actual costs, and because there are many more DBCs than cost pools (hospitals), the DBC prices can easily be set to make everyone look efficient.

Now, as mentioned above, we have used these models to evaluate the 37 potential mergers of pairs of hospitals with a maximal distance of 10 km. Consider first the DEA CRS case. Summary statistics of the overall potential gain E^H and its decomposition into learning effects LE^H, gains after individual learning E^{*H}, harmony

Table 9.5 Farrell input efficiency in alternative frontier models

Model	Mean E	St.dev. E	$\#\{E = 1\}$	Min$\{E\}$
fdh	0.981	0.081	87	0.386
dea-vrs	0.887	0.137	34	0.227
dea-drs	0.865	0.151	32	0.136
dea-irs	0.848	0.133	17	0.227
dea-crs	0.825	0.141	15	0.136
dea-vrs-biascorr	0.829	0.119	0	0.213
dea-vrs-biascorr-c1	0.751	0.115	0	0.191
dea-vrs-biascorr-c2	0.884	0.136	0	0.226
dea-crs-biascorr	0.768	0.125	0	0.124
dea-crs-biascorr-c1	0.722	0.116	0	0.112
dea-crs-biascorr-c2	0.819	0.140	0	0.135
dea-irs-biascorr	0.788	0.118	0	0.205
dea-irs-biascorr-c1	0.740	0.109	0	0.188
dea-irs-biascorr-c2	0.842	0.132	0	0.225
sfa-linear	0.739	0.169	2	0.141
sfa-loglinear	0.819	0.145	0	0.083
sfa-translog	0.831	0.138	0	0.336
sfa-normedlinear-vrs	0.618	0.222	6	0.008
sfa-normedlinear-crs	0.623	0.187	0	0.158
dea-se	0.929	0.080	15	0.594

effects HA^H and size effects SI^H are reported in Table 9.6. With CRS, the size effect SI^H is 1, as there is no gain in resizing with constant returns to scale.

Table 9.6 Potential gains from mergers in DEA-CRS model

Statistics	E	E^*	LE	HA	SI
Average	0.82	0.97	0.84	0.97	1.00
Std.dev.	0.08	0.03	0.07	0.03	0.00
Max	1.00	1.00	1.00	1.00	1.00
Min	0.64	0.88	0.69	0.88	1.00

At an overall scale, we see that the average potential savings in the 37 mergers is 18% ($1 - E = 1 - 0.82$). Indeed, in the detailed results, 17 out of the 37 pairs has an improvement potential of more than 20%, and 32 out of 37 can save more than 10%.

An important part of this potential savings, namely, 16%, is accounted for by learning potentials. Some of the learning potentials can no doubt be activated by benchmarking across hospitals and by developing better incentive schemes relying, for example, on cross hospital relative performance evaluations. However, a merger may also have a positive effect on learning by increasing the scale of process de-

velopment and by serving as a change event where past procedures are re-evaluated and changed.

Ignoring the learning effect, however, the average potential savings is only 3% (the harmony savings), and only 9 out of 37 (or about 24% of the mergers) can generate a savings of more than 5% by simply reallocating resources and tasks. Again, this is theoretically possible without a genuine merger, e.g., by creating interhospital DBC markets, but the reallocation of resources and tasks may be easier inside a merged hospital where problems of asymmetric information and competition over profit shares may be reduced.

These results suggest that the underlying estimated technology is rather linear, i.e., that not only do we have constant returns to scale by assumption but also the output isoquants are rather linear, corresponding to approximately constant rates of substitution between the outputs. This is not entirely surprising; the linear SFA model gives average efficiencies that are quite similar to those of the DEA models, suggesting that the inability to have curved isoquants in this technology does not lead to too much deviation of the actual performance from the estimated best practice frontier.

If we assume instead a VRS technology, the corresponding results are given in Table 9.7.

Table 9.7 Potential gains from mergers in DEA-VRS model

Statistics	E	E^*	LE	HA	SI
Average	1.00	1.12	0.89	0.93	1.20
Std.dev.	0.26	0.22	0.08	0.06	0.21
Max	1.94	1.94	1.00	1.00	1.94
Min	0.72	0.95	0.72	0.83	0.99

In the VRS calculations, several mergers lead to LP problems with no solutions, i.e., several of the merged firms are outside the technology determined by the 97 hospitals. The explanation is that when two hospitals are merged, they will in many cases become very large compared to the existing hospitals (with similar mix of resources and services) and consequently be above the estimated optimal scale size for this mix. In that cases, the existing best practices do not even show that the resulting production plans are feasible.

If we believe firmly in the estimated VRS technology, the interpretation is that it will be impossible to operate hospitals of that size, or, in the case where a solution is found but its score is above 1, that it will be more costly to operate the hospitals jointly than individually. In one specific merger, for example, we found that the estimated net effect was a cost increase of some 19%. This cost increase is the result of three effects. First, since the underlying units are technically inefficient, there is a learning potential of 12%. Also, by reallocating resources and services, some 2% can be saved. The return to scale, however, is rather unfavorable to this merger, corresponding to a cost increase of 38%. The net effect—when correcting

for the fact that these different effects are multiplicative and not additive—is a cost increase of 19%.

Another more likely explanation of these findings is of course that the estimated technology is flawed or at least heavily biased for large units. The bias of the DEA estimated technology is well know, cf. Chap. 6; DEA makes a conservative (cautious) inner approximation of the production possibility set, and in the parts of the production space where observations are more sparse , this bias is larger. Hence, if there are only few large units comparable to the size of a merged one, the best practice model is most likely too pessimistic—and more so the larger the merged hospital. This may explain the rather modest improvement potentials identified in the VRS case. Another indication of bias here may be that the confidence intervals in VRS are broader than in CRS (and IRS), cf. Table 9.5.

Even more fundamentally, one may of course question the VRS assumption using similar reasoning in a theoretical framework: a large entity must be able to do at least as well as any two smaller units into which it could be decomposed because it could simply be (re-) organized as two independently run divisions.

This suggests that we should either use the bias corrected technology or the IRS technology, or both. Alternatively, we could make parallel evaluations using SFA estimate models. The results (summary statistics) of doing this are shown below.

Table 9.8 Potential gains from mergers in DEA-IRS model

Statistics	*E*	*E**	*LE*	*HA*	*SI*
Average	0.82	0.96	0.85	0.96	0.99
Std.dev.	0.08	0.03	0.06	0.03	0.01
Max	1.00	1.00	1.00	1.00	1.00
Min	0.64	0.88	0.71	0.88	0.95

We see in Table 9.8 that the IRS case gives quite similar results on average to the CRS case. The potential savings from individual learning is 15%, from scope (harmony) 4% and from scale (size) 1%. It is interesting to see that even if we acknowledge the possibilities of small units being disadvantaged by their scale, the gains from the merged units operating at larger scales is generally limited, and only about $^1/_4$ of the gains arise from better economies of scope.

Table 9.9 Potential gains from mergers in bias corrected DEA-IRS model

Statistics	*E*	*E**	*LE*	*HA*	*SI*
Average	0.73	0.96	0.76	0.97	0.99
Std.dev.	0.06	0.03	0.05	0.03	0.01
Max	0.89	1.00	0.89	1.00	1.00
Min	0.58	0.89	0.63	0.89	0.95

We see from Table 9.9 that the bias correction increases the overall potential improvement E, but that it is in general the learning affect that picks up almost all the changes in the cost frontier. The scope (harmony) and scale (size) effects are largely unchanged.

In the VRS case, the impact is also mainly in the learning effect, although the negative impact is also a little lower, as expected. Only under special circumstances does boot-strapping eliminate the LP (no-solution) problem. This is illustrated in Table 9.10.

Table 9.10 Potential gains from mergers in bias corrected DEA-VRS model

Statistics	E	E^*	LE	HA	SI
Average	0.89	1.11	0.80	0.93	1.18
Std.dev.	0.23	0.22	0.07	0.06	0.21
Max	1.73	1.94	0.92	1.00	1.94
Min	0.65	0.94	0.65	0.82	0.99

To illustrate the parametric approach, consider the log-linear model. Since we are estimating a cost function and not a production function, the log-linear specification may conflict with the usual convexity properties, i.e., the set T may not be convex. Rather, the log-linear specification allows for gains from specialization as well as potentially genuine global economies of scale. As an aside, these properties can be interesting to allow for. Assuming a truncated normal inefficiency distribution (with underlying mean μ) and normal distributed noise, the maximum likelihood estimates are as shown in Table 9.11 below.

Table 9.11 Loglinear parametric function

Parameter	Coefficient	Std.dev	t-stat
β_0	0.696	0.923	0.754
β_1	0.651	0.141	4.608
β_2	0.204	0.089	2.297
β_3	0.006	0.006	1.048
β_4	0.149	0.103	1.456
β_5	0.000	0.005	0.011
β_6	0.005	0.004	1.198
σ^2	0.476	0.105	4.538
μ	−1.370	0.313	−4.376

It is worthwhile to note that the sum of the beta values is 1.01 suggesting a more or less constant return to scale technology. Of course, several of the parameter are actually not significant which could suggest a re-estimation with fewer cost drivers, but we leave this issue for now.

Using the log-linear specifications we can calculate and decompose the gains from mergers as in Table 9.12 below. The method used is the one described in Sect. 9.3.3 using the implementation described in Sect. 9.4.2 adapted for the cost function, but come in mind out methodological concern in Sect. 9.4.4.

Table 9.12 Potential gains from mergers in loglinear SFA model

Statistics	E	E^*	LE	HA	SI
Average	0.77	1.03	0.75	1.02	1.01
Std.dev.	0.11	0.02	0.11	0.02	0.00
Max	1.00	1.10	0.92	1.09	1.01
Min	0.51	1.01	0.48	1.00	1.01

We see that the log-linear model suggests that the economies of scale are largely neutral to the mergers, as are the economies of scope. In the loglinear specification, even the scope economies mitigate (slightly) against the mergers, corresponding to a cost increase of 2% on average. The log-linear model suggests that the gains arise primarily from learning effects.

The lack of gains from larger scale, and in many cases even losses from the merged units operating at larger scales, has been a consistent finding in the models above. Of course, it must be emphasized also that our analyses build on existing practices only. If a new merger leads to new facilities and new organizations that have not been implemented in other hospitals in the data set, the estimated models cannot capture the potential savings that these improvements may generate. This would require a much more detailed organizational and engineering approach. A network approach, cf. below, could potentially be developed in this direction. Specifically, if one can define hospital processes and allocate not only activities but also costs to these processes, then it is possible to create new pseudo-observations by constructing new combinations of old processes.

We also note that the Spearman correlation between the individual efficiencies calculated in the log-linear model and the DEA-IRS model is 0.62, while it is 0.68 in the DEA-IRS bias corrected model and 0.68 in the DEA-CRS model. In general, then, these models suggest correlation but not perfect agreement in the individual evaluations. This illustrates a point that was also emphasized above. The models analyzed here cannot directly be used as authoritative cost models for Dutch hospitals. In the analysis of specific merger cases, it is important to develop good underlying production and/or cost models of the technology in place. It is likely, however, that even after such efforts, there may be several reasonable model candidates. The best approach in this situation may be to evaluate the merger gains in the different models—as we have done here—and to look at the results as interval estimates established in this way.

9.6 Practical application: Mergers of Norwegian DSOs

The Norwegian regulator of electricity networks, The Norwegian Water Resources and Energy Directorate (NVE), has adopted the above framework in their determination of the conditions for mergers among concession holders.

Their procedures compares the sum of the cost norms to each of the involved firms with the cost norms that result if they are treated as a single, merged entity. This difference, which by the DEA based cost norm model used is in fact equivalent to the harmony effect (and is so called), is then used to correct the cost norm calculated for the merged firm. Specifically, the net present value of the harmony effect over 10 years are calculated and paid as a windfall gain to the merged firm. In effect, this means that the extra saving potential measured by the harmony effect can be kept by the firms for the first 10 years. Hereafter, the savings must be transferred to the end users.

From a regulatory point of view, this approach makes sense. A possible drawback of a many regulatory systems is that they tend to freeze structural developments, i.e., changes in the industrial structure that can lower costs. With a convex cost norm based revenue cap, the firms will always be better off in terms of allowed costs before a merger than after a merger. Therefore, the firms must be compensated in order to give them incentives to reduce costs via mergers.

9.7 Controllability, transferability, and ex post efficiency

In the estimates of potential merger gains above, we have assumed that all inputs and outputs can be redistributed in the merged entity H. In many cases, this assumption is too restrictive. At least from a short term perspective, some dimensions are easier to change and reallocate than others. It may, for example, be easier to reduce the labor input than the capital input, which is largely based on sunk investments. Also, some services may have to be provided on location and can therefore not be transferred to another firm located elsewhere. In a hospital setting, for example, it may be possible to transfer IT, accounting and HR to another location, but the provision of emergency room services cannot easily be relocated. Lastly, some variables in actual models typically describe context rather than choice variables, and they are therefore not transferable. Population density, education level, and age distribution, for example, have limited transferability.

We will now show how to calculate potential gains when only some of the inputs and outputs can be *adjusted* and *transferred* among the members of the new merged entity. First, we consider the relatively straightforward case with restricted controllability of inputs and outputs, and then we extend the model by introducing restrictions on the transferability of some of the resources and services.

As discussed already in Sects. 2.4.1 and 5.3, it is common to account for the *non-discretionary* character of some dimensions by only looking for improvements in the other directions. Assume that we can split the inputs x and outputs y into two types,

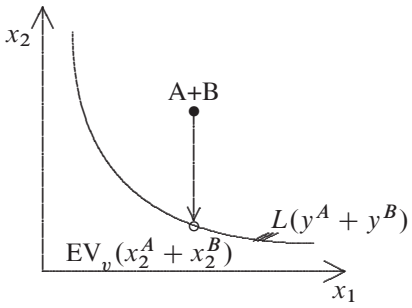


Fig. 9.5 Merger evaluation with restricted controllability

$x = (x_v, x_f)$ and $y = (y_v, y_f)$, corresponding to the variable (controllable) v and fixed f (non-controllable) dimensions. Focusing on input reductions, we would now look for the largest reduction of all controllable inputs that, together with the fixed inputs, allows the merged firm to produce the given outputs.

To illustrate, consider the case in Fig. 9.5 below. We assume that the integrated firm $A + B$ is using two inputs, e.g., doctors and nurses, in its production. Assume also that the first input cannot be adjusted but that the second can. In an application, for example, this may reflect a case where the doctors have more bargaining power or are on more rigid contracts than the nurses. The efficiency of the merged unit will therefore be measured by the possible reduction in the second input E_v alone.

To formalize the sub-vector approach to merger analysis, we measure the efficiency of firm i as

$$E_v^i = \min\{E \in \mathbb{R}_+ \mid (Ex_v^i, x_f, y^i) \in T\}.$$

Likewise, the potential gross gain from a horizontal merger of the H firms is given by

$$E_v^H = \min\left\{ E \in \mathbb{R}_+ \mid \left(E \sum_{k\in H} x_v^k, \sum_{k\in H} x_f^k, \sum_{k\in H} y^k\right) \in T \right\}$$

i.e., E_v^H is the maximal proportional reduction in the variable (discretionary) inputs that, together with the fixed (nondiscretionary) resources, allows the production of the aggregated output profile $\sum_{k\in H} y^k$. If $E_v^H < 1$, we can save the proportion $(1 - E_v^H)$ of the variable inputs by merging. If $E_v^H > 1$, the merger is costly and requires that the total usage of the variable resources be increased.

As previously observed, we may also filter out the effects of individual inefficiencies by determining the *adjusted overall gains* in the direction of the variable inputs as

$$E_v^{*H} = \min\left\{ E \in \mathbb{R}_+ \mid \left(E \sum_{k\in H} E_v^k x_v^k, \sum_{k\in H} x_f^k, \sum_{k\in H} y^k\right) \in T \right\},$$

and we can define the *learning effect* as the ratio

$$LE_v^H = \frac{E_v^H}{E_v^{*H}}.$$

Likewise, we can calculate the *harmony effect* as

$$HA_v^H = \min\left\{ HA \in \mathbb{R}_+ \mid \left(HA\,\frac{1}{H}\sum_{k\in H} E_v^k x_v^k, \frac{1}{H}\sum_{k\in H} x_f^k, \frac{1}{H}\sum_{k\in H} y^k\right) \in T \right\},$$

and the *size effect* as

$$SI_v^H = \min\left\{SI \in \mathbb{R}_+ \mid \left(SI \cdot HA_v^H \sum_{k\in H} E_v^k x_v^k, \sum_{k\in H} x_f^k, \sum_{k\in H} y^k\right) \in T\right\}.$$

The interpretations and organizational implications of these scores are the same as previously explained, except that they are now defined in terms of the savings of only the controllable inputs, calculated conditional on the given levels of the noncontrollable inputs. Thus, for example, rescaling is advantageous, $SI_v^H < 1$, if we have economies of scale in (x_v, y) for given x_f, and costly, $SI_v^H > 1$, if the returns to scale do not favor larger firms for the given values of the fixed inputs. Using the above definitions, we once again get a decomposition

$$E_v^H = LE_v^H \cdot HA_v^H \cdot SI_v^H.$$

This corresponds to a decomposition of the basic merger index E_v^H into a technical efficiency index TE_v^H, a harmony index HA_v^H, and a size index S_v^H.

So far, we have dealt with the possibility that only some of the variables are discretionary within a given time horizon. Another obstacle to the reallocation among firms may be *nontransferable (local, l)* resources and services as opposed to *transferable (global, g)* ones.

To illustrate this, consider Fig. 9.6. We have two service providers, and to simplify the interpretation, they produce the same globally transferable outputs. Also, we assume that they use the same technology. Now, if input 2, e.g., nurses, is transferable and input 1, e.g., doctors is not, we could move some x_2 (nurses) from B, where they have a rather low marginal value compared to x_1 (doctors), to A, where their marginal value is higher. In the end locations, B will have its output reduced (from say y^B to y^A), but A will have its output increased from y^A to some level $y^* > y^B$. The net result is therefore positive even though we cannot reallocate the factors as easily as in the usual harmony calculations.

Allowing for possibly restricted transferability and possibly restricted controllability, we get a 2×2 taxonomy of all inputs and outputs: They may be lf (local and fixed, e.g., buildings), lv (local and variable, e.g., cleaning personnel), gf (global and fixed, e.g., specialized measurement equipment), and gv (global and variable, e.g., different types of specialists). To simplify notation, we can in the usual way indicate vectors of such variables by suppressing the subscripts. We will for example refer to the l variables as the combination of the lf and lv variables, and to the f factors as the combination lf and gf factors.

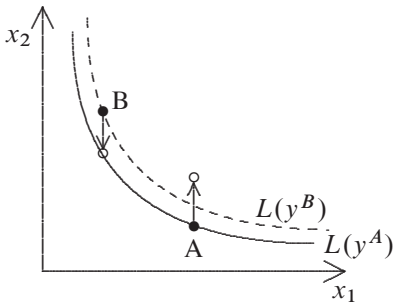


Fig. 9.6 Restricted transferability

In such a setting, it is more complicated to measure the possible gains from a reallocation of resources and services among the otherwise autonomous firms in H.

A harmony measure HA in this case could be calculated in the following way. We consider what can be saved—after individual learning—by mere reallocations of the global inputs and outputs among the firms in H. Let the new input and output combinations after reallocation be (x^{*k}, y^{*k}) for $k \in H$. We then look for such new production plans for all the firms involved that maximize the aggregate savings. This leads to the *pure reallocation problem*:

$$\min_{HA,(x^{*k},y^{*k})_{k\in H}} HA$$

$$\begin{aligned}
\text{s.t.} \quad HA \sum_{k\in H} E_v^k x_v^k &\geq \sum_{k\in H} x_v^{*k} && : \text{ variable factors are reduced}\\
x_l^{*k} &\leq x_l^k \quad (k \in H) && : \text{ local factors are saved locally}\\
\sum_{k\in H} x_{gf}^{*k} &\leq \sum_{k\in H} x_{gf}^k && : \text{ global fixed factors are not reduced}\\
y_l^{*k} &\geq y_l^k \quad (k \in H) && : \text{ local services are produced on-site}\\
\sum_{k\in H} y_g^{*k} &\geq \sum_{k\in H} y_g^k && : \text{ global services can be prod. off-site}\\
(x^{*k}, y^{*k}) &\in T \quad (k \in H) && : \text{ all plans are technically feasible.}
\end{aligned}$$

The choice variables in this program are the contraction factor HA and the new input and output combinations $(x^{*k}, y^{*k})_{k\in H}$. Because the original adjusted productions $(E_v^k x_v^k, x_f^k, y_l^k, y_g^k)$ for $k \in H$ satisfy all the constraints, we always have $HA \leq 1$, corresponding to a potential savings.

When all resources and services are global and variable, the above program leads to the reallocation problem we discussed above as a rationale for the harmony efficiency.

A potential drawback of most of the models in the literature, as well as the models presented above, is that they assume that the reallocations take place at the frontier, i.e., they presume *ex-post efficiency*. This means that all firms are assumed to adapt

to the best practices before reallocation occurs. Although competition may work to drive out inefficient firms, one can see it as naive to assume technical efficiency up front. Empirical studies have shown that inefficiency is a persistent phenomenon in many industries, and one can even find theoretical economic rationales for maintaining some technical inefficiency. It may, for example, help to compensate employees by making their job more attractive, or it may improve the strategic interactions with other firms on the market by showing excess capacity. Alternatively, the idea of reallocations on the frontier presumes that all entities are profit maximizing, which is certainly not the case in many of the sectors for which performance evaluation is relevant, e.g., in the health care sector, where immediate short term adaption to best practices may not be realistic.

In the discussion of more extended reallocation models it has been investigated how to avoid the assumption of ex-post efficiency. That is, one can assume that if firms have been inefficient in the past, it is likely that they will continue to be so in the future. One can also assume that the level of future inefficiency may depend on the extent to which the firms have to change the scale and scope of their operations.

9.8 Summary

In this chapter, we discussed how to measure a priori the potential gains from restructuring a set of firms. We have developed an overall measure of the potential gains from mergers. The potential gains are simply $1 - E^H$, where E^H is the efficiency of the aggregated firm using the sum of the original inputs of the firms in H to produce the sum of the original outputs from the H firms.

We also decomposed the overall efficiency E^H into the possible learning effect, harmony effect and size effect. The relevance of this decomposition stems from the fact that some of the gains can be realized using alternatives to full scale mergers. The learning potentials may at least partially be captured by sharing information on best practices, and the harmony effects can be realized by reallocating resources and tasks among the firms in H. Such alternative measures are relevant because mergers in many industries come with drawbacks as well. For example, the integration of different organizational cultures may be cumbersome, and from a regulatory point of view, mergers tend to lower the competitive pressure in the industry.

We have extended the basic measures in various ways. In particular, we have discussed how to model the restricted controllability and restricted transferability of resources and services.

In addition to the basic measures and decompositions, we have shown how to implement the methods in R, and we have given some highlights from an application to Dutch hospitals. Other applications are mentioned in the bibliographical notes below.

It is worthwhile to compare the ideas of this chapter with the idea of allocative efficiency discussed already in Chap. 2. Allocative efficiency is typically defined as cost efficiency divided by technical efficiency. It therefore measures what can

be gained by adapting to given prices in a complete and perfect market. Therefore, allocative efficiency effectively ignores the *matching issues* in a finite economy. The merger analysis explicitly takes such complications into account by measuring the reallocation gains among a finite set of firms in cases where no market prices exist.

9.9 Bibliographic notes

The idea of structural and allocative efficiencies was introduced early in the quantitative literature. Farrell (1957) defines structural efficiency as the extent to which an industry keeps up with the performance of its own best firms, and he suggests measuring it by comparing the horizontal aggregation of the industry's firms with the frontier constructed from its individual firms. A related approach is the average firm approach suggested by Førsund and Hjalmarsson (1979). In this approach, the structural efficiency is estimated by taking the average of each type of input and each type of output and then measuring the distance from the associated average firm to the frontier. This idea is clearly closely related to the way we measure the gross potential gains from a merger of all firms—or the harmony effect of a merger of all firms. Lewin and Morey (1981) discuss the decomposition of inefficiency in a hierarchical organization into what can be attributed to inefficiencies in the production firms with given resources and the misallocation of resources among the firms at different levels of the organization. Färe and Grosskopf (2000) further develop the idea of modeling efficiency in different types of network structures

The basic approach of this chapter was first suggested in Bogetoft and Wang (2005) and Bogetoft et al (2003). There, we studied the potential gains from mergers in different industries, in particular, Danish agricultural extension service and Danish forest organizations. The reallocations were restricted to taking place among geographical neighbors. Moreover, the gains were decomposed into learning, harmony and size effects, and the corresponding organizational changes were identified. The somewhat unusual term harmony is motivated by the idea that some factor combinations are more harmonious or in balance than others. When we developed this idea, we were working on a problem of harmony requirements in Danish agriculture, and the term later got used in several applications and in some regulation systems, which is why we still use it. We introduced restricted controllability merger analyses in an analysis of Norwegian DSO mergers, Bogetoft and Gammeltvedt (2006), and we discussed the difference between local and global resources and obligations in Bogetoft and Katona (2008). Decompositions of merger gains using a directional distance function approach, which, as mentioned in the text, leads to additive instead of multiplicative decompositions, are given in Bogetoft and Katona (2008).

We have focused in this paper on horizontal mergers, but a parallel treatment of the case of vertical integration of firms in a supply chains can be found in Bogetoft and Katona (2008). We have also discussed additive decompositions resulting from the use of directional distance functions.

The Dutch hospital analyses are further developed in Bogetoft and Katona (2008), and a recent application to Danish hospitals is presented in Bogetoft et al (2010).

Studies applying the more extended reallocation models, where firms may not adjust to best practice and where the level of ex post efficiency may depend on the scale and scope of the reallocations are discussed in Andersen and Bogetoft (2007) and Bogetoft et al (2007a).

Chapter 10
Regulation and Contracting

10.1 Introduction

One of the more prominent applications of state-of-the-art benchmarking is in the regulation of natural monopolies in general and electricity and gas networks, in particular. Benchmarking studies applied to inform such regulation has considerable economic impact on firms and consumers alike.

In this chapter, we will describe some classical regulatory packages and explain the role of benchmarking in these scenarios. Also, we will illustrate some of the models that have been developed in a selection of countries. Regulation and benchmarking theory have traditionally been two separate, although mutually supportive, economic disciplines, but during the last 15 years, there has also been some novel work done to integrate the two. We will cover some key results and an application towards the end of this chapter.

10.2 Classical regulatory packages

Large infrastructure industries like the networks to distribute electricity and gas, commonly referred to as Distribution System Operators DSOs, constitute natural monopolies; there is considerable fixed cost and relatively low marginal costs. This leads to market failure. Moreover, such networks are generally given licenses to operate as legal monopolies. Monopolies have limited incentives to reduce costs, and will tend to under-produce and overcharge the services provided since they are not subject to the disciplining force of the market. For electricity distribution, the monopoly characteristic is accentuated by the fact that there are no close substitutes for the offered services and that demand is relatively inelastic.

Most countries therefore empower regulators to act as a proxy purchaser of the services, imposing constraints on the prices and the modalities of the production. The regulator is usually affiliated with the national competition authority. One of

P. Bogetoft and L. Otto, *Benchmarking with DEA, SFA, and R*, International Series in Operations Research & Management Science 157, DOI 10.1007/978-1-4419-7961-2_10,

the instruments used in the regulation is benchmarking, which is facilitated by the existence of different networks covering different areas that can be compared or, in some cases, by international comparisons of such firms.

In modern economic theory, the regulatory problem is expressed as a game between a principal (the regulator) and a number of agents (the regulated firms). The regulation problem is basically one of controlling one or more firms that have superior information about their technology and their cost reducing efforts as compared to the regulator. The availability and access to information is a key issue in the regulatory game. With perfect access to information, the regulator can impose socially optimal price and service quality.

The regulatory toolbox contains numerous more or less ingenious solutions to the regulator's problem. To illustrate, we will distinguish four approaches

- Cost-recovery regimes (cost of service, cost-plus, rate of return),
- Fixed price (revenue) regimes (price-cap, revenue cap, RPI-X),
- Yardstick regimes, and
- Franchise auction regimes.

10.2.1 Cost-recovery regimes

Taking for granted the cost information supplied by the agents, the regulator may choose to fully reimburse the reported costs, often padded with some fixed mark-up factor. To illustrate, the reimbursement in a given period t for firm k may be determined as

$$R^k(t) = C^k_{OpEx}(t) + D^k(t) + (r + \delta)K^k(t)$$

where C^k_{OpEx} is the operating expenses , K^k is the capital (rate base), D^k is the depreciation reflecting capital usage, r is the interest rate reflecting the credit costs of investments with similar risks and δ is a mark-up.

Unless subject to costly information verification (regulatory administration), the approach results in poor performance with skewed investment incentives (no investment risk, yet fixed return on investment), perverse efficiency incentives (loss of revenue when reducing costs) and insufficient managerial effort.

In reality, such schemes have involved considerable regulatory administration, in an attempt to avoid imprudent or unreasonable operating expenditures and investments to enter the compensation and rate base. Some benchmarking approaches use, for example, a few key performance indicators KPIs from similar firms. However, even with large investments in information gathering, the information asymmetry and the burden of proof in this regime rest on the regulator, and there are reasons to doubt their ability to induce efficiency.

Regulatory authorities worldwide are gradually abandoning these regimes because they are administratively costly and technologically inadequate, especially in

the United Kingdom (U.K.) and the United States of America (USA), where the approach has been heavily used.

Cost recovery is often organized as negotiation and consultation based regimes. Whether rate reviews are initiated by complaints or are planned, reviews are often done as individual consultations. In contrast to the methods below, where a joint framework is used to evaluate all DSOs, the consultations are typically case-specific and they rely more on negotiations than on a comprehensive model estimation for the entire sector.

An idea is to combine negotiations with systematic investigations and benchmarking in such a way as to limit the negotiation space, In this way, the negotiations become more structured. Such restrained negotiations have been proposed in the Netherlands for the regulation of hospitals; the idea is that the regulator uses benchmarking to constrain acceptable outcomes but leaves negotiations to industry partners.

10.2.2 Fixed price regimes (price-cap, revenue cap, CPI-X)

In response to the problems of the cost-recovery regime, several countries have moved to more high-powered regimes. These regimes typically allow the regulated DSOs to retain any realized efficiency gains. In the price-cap regime, the regulator caps the allowable price or revenue for each DSO for a pre-determined regulatory period, typically 4-5 years. Based on the review period, a model of probable cost developments is developed to fix the revenue or price basket. The base model is usually quite simple, involving a predicted productivity development per year x plus, perhaps, individual requirements on DSOs, x^k, to reflect the level of historical costs and thereby the need to catch-up to best practice. The resulting allowed development in the revenue for DSO k is then

$$R^k(t) = C^k(0)(1 - x - x^k)^t, \qquad t = 1, \ldots, T$$

where $R^k(t)$ is the revenue in period t and $C^k(0)$ is the cost of DSO k in period 0. Note that x is used here not as input but as an efficiency requirement; this is in accordance with the standards in regulations where the above model is often referred to as CPI-x to reflect that there are adjustments for price developments and productivity requirements. There are, of course, many modifications to this model. Thus, there will typically be adjustments for changes in the volume supplied and for general changes in the cost level due to inflation. We have already seen one such example in Sect. 2.7, and will show another example from Germany below.

The crucial feature of the fixed price regime is that there is a fixed (performance independent) payment. This means that, to maximize profit, the DSO will minimize costs. This is key to the incentive provision.

Another important feature is the fixation of payments during a regulatory period and the consequent regulatory lag in updating productivity development. The last

feature is often emphasized by calling such schemes *ex ante regulation* as illustrated in Fig. 10.1 below. Before a regulatory period starts, the regulator uses historical data from a review period to estimate x and x^k, and then commits to these values for the regulatory period of T years. At the end of the regulatory period, new estimations of x and x^k are made to set the revenue conditions for the next regulatory period.

Fig. 10.1 Ex ante regulation

The idea of price or revenue fixation is simple but in practice the cap is regularly reset, in hindsight, to the realized profits in the previous period. This limits the efficiency incentives. Also, the initial caps have to strike a careful balance between informational rents, incentives for restructuring and the bankruptcy risks. Further, the price or revenue cap is usually linked to the consumer price index (CPI) or the retail price index (RPI) as a measure of inflation. Therefore, in spite of its conceptual simplicity, the challenges of fixing the initial caps, the periodicity of review and the determination of the X-factor make this regulation a non-trivial exercise for the regulator. In particular, since initial windfall profits are retained by the industry and dynamic risks are passed on to consumers, there is a potential risk of regulatory capture by consumer or industry organizations.

For now, however, the most important feature is that the price fixation regimes generally involve some systematic benchmarking exercise, often based on DEA and SFA, to guide the choice of individual requirements x^k and the general requirement x.

The general requirement x is often set by using a Malmquist-like analysis of productivity developments over the years prior to the regulatory period. Thus, if the analysis of past frontier shifts suggests that even the best are able to reduce costs by 2 % per year, the regulator has a strong case to set x close to 2%.

Individual requirements x^k are typically linked to the individual efficiencies of the DSOs in the last period prior to the regulatory period. There are no general rules used by regulators to transform a Farrell efficiency E^k to an individual requirement x^k, except that the smaller E^k is, the larger x^k is. Some countries require the DSOs to catch-up very quickly. In the first Danish regulation of electricity networks, for example, the electricity producers were required to eliminate the inefficiency in just 1 year. Others, like the Netherlands, used one regulatory period of 3-5 years. Germany aims to have eliminated the individual efficiency differences in two periods, i.e. 10 years, while Norway, a pioneer in the use of incentive-based regulation, allowed for an even longer period of time in the initial implementation of a revenue cap system. It is clear that the analyses of historical catch-up values can guide this

decision. There is also a considerable element of negotiation in the rules that are applied. Moreover, it is difficult to compare these requirements across countries. A cautiousness principle would suggest that the requirements will depend on the quality of data and the benchmarking model. Also, a controllability principle would suggest that it should depend on the elements that are benchmarked. In particular, it is important if it is Opex (operating expenses) or Totex (= Opex+Capex) that are being benchmarked and that become subject to efficiency improvement requirements.

In Denmark, for example, the first model from 2000 had very rigorous requirements on Opex - but still allowed new capital evaluations (opening statements), which lead to increased Capex allowances. On average, the companies only used 80-85% of the revenue caps. This suggests that the regulation may not have been as rigorous as it looked (with immediate catch-up requirement in a linear model), nor was the importance of consumer preferences in the many cooperatively-owned distribution companies foreseen. Either way, this led to immense accumulated reserves by the end of 2003. In return, this meant that adjustments in the regulation could have only limited impact since the DSOs could always draw on past revenue cap reserves. The regulation was, therefore, abandoned at the end of 2003 and a new regulation was later established.

We will give some more detailed illustrations of some of the steps in regulatory benchmarking for revenue cap regulation in Sect. 10.3 below, where we discuss the recently developed German benchmarking model.

10.2.3 Yardstick regimes

The idea behind yardstick regimes is to mimic the market as closely as possible by using real observations to estimate the real cost function in each period rather than relying on ex ante predicted cost functions. Thus, for example, in its simplest form, the allowed revenue for DSO k in period t would be set ex post and determined by the costs in the same period of other firms $h = 1, \ldots, k-1, k+1, \ldots, K$ operating under similar conditions

$$R^k(t) = \frac{1}{K-1} \sum_{h \neq k} C^h(t), \qquad t = 1, 2, \ldots$$

Observe that this is the revenue the firm could charge in a competitive environment.

Of course, if the DSOs are delivering different services under different contextual constraints, the above revenue cap is not directly applicable. Instead, we use benchmarking to account for these differences. Also, one can argue that the average is just one of many ways to aggregate the performance of the other firms. One alternative is to use best practice realized performance, i.e.

$$R^k(t) = \min\{C^h(t) \mid h = 1, \ldots, k-1, k+1, \ldots, K\}, \qquad t = 1, 2, \ldots$$

The yardstick regime is attractive in the sense that the revenue of a given DSO is not determined by its own cost but by the performance of the other DSOs. This fixed price feature makes the firm a residual claimant, as in the price fixation regime, and this is the key incentive property.

Another advantage of yardstick competition is that the productivity development is observed rather than predicted. This provides insurance for the DSOs, and it limits the information rents. It is accomplished by setting the revenue *ex-post*, i.e. after each period. This is illustrated in Fig. 10.2. The allowed costs in period t is only set after period t. Exogenous and dynamic risks will directly affect the costs in the industry, lifting the yardstick. Innovation and technical progress will tend to lower the yardstick. Thus, the regime endogenizes the ubiquitous x factor and caps the regulatory discretion at the same time.

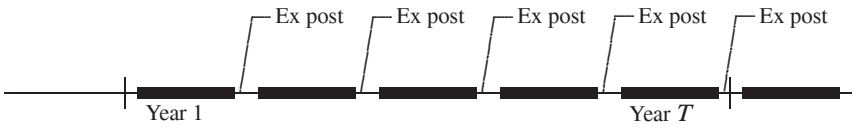


Fig. 10.2 Ex post regulation

Despite its theoretical merits, the pure approach of only considering the observed cost in each period is linked to some risks in implementation. First, a set of comparators with correlated operating conditions must be established. Second, if the comparators are few and under similar regulation, there is risk of collusion. Finally, a yardstick system that is not preceded by a transient period of asset revaluation or franchise bidding will face problems with sunk costs and/or bankruptcy. A crucial question, in terms of yardsticks in electricity distribution, is, therefore, how to preserve the competitive properties while assuring universal and continuous service.

In Sect. 10.4 below we will expand on the advantages of the yardstick idea and we will show how to cope with cases of imperfectly correlated costs and variations in output levels and mix by using DEA.

From the point of view of benchmarking, the yardstick regime requires the same model types as price fixation regimes, only now benchmarking has to take place more often, typically annually. A DEA-based yardstick scheme was introduced in Norway 2007 and will be discussed later. Also, the Dutch regulation has yardstick features.

10.2.4 Franchise auctions

A simple means to elicit accurate cost information while assuring participation is to arrange franchise auctions. The idea is to award the delivery rights and obligations based on an auction among qualified bidders. Thus, for example, if each of K bidders for a project demands $B^h, h = 1, \ldots, K$, we may award it to the bidder k with the lowest bid $B^k = \min_h B^h$ and compensate him

$$R^k(t) = \min_{h \neq k} B^h(t), \qquad t = 1, 2, \ldots$$

Here, we have used what is often called a second-price approach, and we see that it resembles a yardstick regime. We do, however, use bids rather than realized costs in the auction scenario. One can extend this scenario to situations with heterogenous bids by using, for example, DEA based auctions to cope with differences in the projects offered in a one-shot procurement setting.

The second-price franchise auction regime conserves the simplicity of the fixed-price regimes but limits the informational rent. It also offers perfect adjustment to heterogeneity, since prices may vary across franchises. The problems for limited markets with high concentration are that bidding may be collusive, that excessive informational rents may be extracted and that competition may be hampered by asymmetric information among incumbents and entrants. Even under more favorable circumstances, the problems of bidding parity, asset transition and investment incentives must still be addressed, and the use of the franchising instrument in, for example, electricity distribution is likely to be used sparingly in Europe in the near future and then primarily for spatial and/or technical service extensions.

10.2.5 Applications

Table 10.1 below gives a summary of the regulations used for electricity DSOs in 15 European countries. Dynamically, the progression seems to be from a more heavy-handed cost recovery regime, over a model-based price fixation towards a high-powered market-based yardstick regime.

Most countries rely on some revenue cap model and have derived general productivity and individual inefficiencies using benchmarking tools like DEA and SFA.

We see how some countries, like Sweden and Spain, have chosen to rely on technical engineering norms, sometimes referred to as ideal nets, in an attempt to identify not only best practice but absolute technological possibilities.

Table 10.1 Some European regulations of electricity DSOs

Code	Country	Regulation	Benchmark
AT	Austria	Revenue cap	DEA-SFA, best-off
BE	Belgium	Revenue cap	DEA
CH	Switzerland	Cost recovery	Ad hoc
DE	Germany	Revenue cap	DEA-SFA best-off
DK	Denmark	Revenue cap	COLS-MOLS
ES	Spain	Revenue cap	Engineering
FI	Finland	Revenue cap	DEA w. SFA back-up
FR	France	Cost recovery	Ad hoc
GB	Great Britain	Revenue cap	COLS and Ad hoc
GR	Greece	Cost recovery	Ad hoc
HU	Hungary	Price cap	Ad hoc
IRL	Ireland	Price cap	Ad hoc
NL	Netherlands	Yardstick comp	DEA-OLS-MOLS
NO	Norway	Yardstick comp	DEA
SE	Sverige	Revenue cap	Engineering and DEA

10.3 Practical application: DSO regulation in Germany

In this section we will discuss the regulation of electrical DSOs in Germany. We will explain some of processes leading to the regulation and go through some highlights of the benchmarking models used.

10.3.1 Towards a modern benchmark based regulation

In 2005, it was decided to introduce new regulation of German electricity and gas DSOs. Here, we will focus on regulation of electricity, but we note that gas regulation and models are rather similar. Previously, regulation occurred solely through competition law, and there was no regulator. With the new Electricity Act (EnWG), effective July 13, 2005, it was decided that "Regulation should be based on the *costs of an efficient and structurally comparable operator* and provide incentives based on efficiency targets that are *feasible and surpassable*."

The enactment of the Electricity Act marked the start of an intense and ambitious development process by the regulatory authority, the Federal Network Agency, Bundesnetzagentur (BNetzA). BNetzA performs tasks and executes power, which under the EnWG has not been assigned to the state regulatory authorities. The state regulatory authorities are responsible for regulating power supply companies with fewer than 100,000 customers connected to their electricity or gas networks and whose grids do not extend beyond state borders. In practice, the BNetzA approach has a significant impact on the regulation of the DSOs under state regulation.

Through several development projects and a series of consultations with industry on the principles, BNetzA developed a specific proposal for how to implement the Electricity Act. As one of several consulting groups, we undertook a series of full-scale trial estimations of different model specifications. DEA and SFA models were developed based on more than 800 DSOs in both sectors. This served several purposes, some of which were to train the regulatory personnel in benchmarking methodology, to guide future data collection, to define a detailed implementation plan and to facilitate an informed discussion with industry members.

The final proposal and detailed implementation plan by the regulator was largely transformed into the Ordinance that now provides specific guidelines for German regulation of electricity.

During 2008, we developed a new set of results to implement the Ordinance. Some highlights from this work are provided below. The new regulation became effective in 2009 for the 200 DSO under federal regulation. Smaller DSOs, with no more than 30,000 customers connected directly or indirectly to their electricity distribution system, could, instead of efficiency benchmarking to establish efficiency levels, take part in a simplified procedure. The efficiency level in the first regulatory period for participants in the simplified procedure is 87.5 percent. From the second regulatory period, the efficiency level for these DSOs is the weighted average of all efficiency levels established in nationwide efficiency benchmarking.

The regulation is currently in place and working, although there are still some aspects that are being tested in the court system by different operators.

From an international perspective, the German experience is remarkable because of the large number of DSOs, the abundance of data, as illustrated by the presence of about 250 variables for each DSO, and by the speed and efficiency with which new regulation was established. Most other regulators have used a considerably longer period of time to undertake considerably less ambitious prototyping and full scale implementation.

10.3.2 Revenue cap formula

The German regulation is basically a revenue cap regulation. Each regulatory period is 5 years and the content of the first two regulatory periods have been detailed, giving the DSO more long-term forecasts on which to act.

It is also a Totex based regulation, i.e. both operating expenses (Opex) and capital cost expenses (Capex) are subject to regulation. Capital costs are based on either book values or standardized costs using replacement values and constant annuity calculations of yearly cost using life times of different asset groups.

The revenue cap of an individual DSO k in the German regulation in year t is determined by the formula

$$R^k(t) = C^k_{nc}(t) + (C^k_{tnc}(0) + (1 - V(t))C^k_c(0))(\frac{RPI(t)}{RPI(0)} - x(t))ExFa(t) + Q(t)$$

where C_{nc} is the cost share that cannot be controlled on a lasting basis (statutory approval and compensation obligations, concession fees, operating taxes etc.), C_{tnc} is the cost share that cannot be controlled on a temporary basis (essentially the efficient cost level found as the total costs multiplied by the efficiency level, C_c are the controllable costs, $V(t)$ is a distribution factor for reducing inefficiencies (initially set to remove incumbent inefficiency after two regulatory periods, i.e. 10 years), RPI(t) is the retail price index in year t, $RPI(0)$ is the retail price index in year 0, and $x(t)$ is the general productivity development from year 0 to year t reflecting the cumulative change in the general sectoral productivity factor for year t of the particular regulatory period relative to the first year of the regulatory period. Also, ExFa is an expansion factor reflecting the increase in service provision in year t compared to year 0 and determined as

$$ExFa_j^k(t) = 1 + max(\frac{L_j^k(t) - L_j^k(0)}{L_j^k(0)}, 0)$$

where $L_j(t)$ is the volume of load at level j in year t of the particular regulatory period. The expansion factor for the entire network is the weighted average of all network levels. Lastly, Q(t) is the increase or decrease in the revenue cap from quality considerations. Revenue caps may have amounts added to or deducted from them if operators diverge from required system reliability or efficiency indicators (quality element). The quality element is left to the discretion of the regulator.

10.3.3 Benchmarking requirements

From a benchmarking perspective, the regulation is remarkable for being explicit with respect to a series of technical aspects such as cost drivers, estimation techniques, return to scale and outlier criteria.

The Ordinance is specific about a *minimal set of cost drivers*. Cost drivers such as connections, areas, circuit length, and peak flow, were obligatory. Of course, this leaves a series of available alternatives even within these groups and it does not exclude cost drivers covering other aspects of the service provision.

The German incentive regulation is also explicit as to which *estimation techniques* to use in benchmarking electricity and gas DSOs and how to combine the results of multiple models. According to Section 12 of the Ordinance, the efficiency level for a given DSO is determined as the maximum of four efficiency scores, $E_{DEA}(B)$, $E_{DEA}(S)$, $E_{SFA}(B)$, and $E_{SFA}(S)$, where E_{DEA} is the Farrell efficiency, calculated with a NDRS-DEA model, E_{SFA} is the Farrell input efficiency, calculated using a SFA model, and the argument B denotes book value and S standardized capital costs. As such, the regulation takes a cautious approach and biases the decision in favor of the DSOs in case of estimation risk. Entities demonstrating particularly low efficiency are given the minimum level of 60 percent. In summary, the efficiency of DSO k is calculated using this equation

$$\max\{E^k_{DEA}(B), E^k_{DEA}(S), E^k_{SFA}(B), E^k_{SFA}(S), 0.6\}$$

It is worthwhile noting that the Ordinance does not prescribe any bias correction for the DEA scores, nor does it rely on confidence intervals for the scores, as they could be calculated in both the DEA model (via boot-strapping) or in the SFA model (directly from the estimated variances of the noise and inefficiency terms).

The Ordinance is also specific about *how to identify outliers*. Indeed, it prescribes two outlier criteria to be tested for each DSO, and if any of them is fulfilled, the DSO cannot be allowed to affect the efficiency of the other DSOs. The two criteria can be formalized in the following ways. Let $K^* = \{1, \ldots, K\}$ be the DSOs is the data set, and k be a potential outlier. Also, let, $E(h, K^*)$ be the efficiency of h when all DSO are used to estimate the technology and let $E(h, K^* \setminus k)$ be the efficiency when DSO k does not enter the estimation.

The *first outlier criterion* is that a single DSO should not have too large of an impact on the average efficiency. We can evaluate the impact on the average efficiency by considering

$$\frac{\sum_{h \in K^* \setminus k} (E(h, K^* \setminus k) - 1)^2}{\sum_{h \in K^* \setminus k} (E(h, K^*) - 1)^2}$$

The test compares the average efficiency of the other operators when k cannot affect the technology as compared to the average efficiency of the other DSOs when the k is allowed to impact the evaluations. Since $E(h, K^* \setminus k) \geq E(h, K^*)$, this ratio is always less than or equal to 1, and the smaller the ratio is, the larger the impact of k, i.e. small values of the ratio will be an indication that k is an outlier. The asymptotic distribution of the ratio is $F(K - 1, K - 1)$.

The *second outlier criterion* is that no DSO k will be extremely super-efficient in the sense that

$$E(i, K^* \setminus k) > q(0.75) + 1.5(q(0.75) - q(0.25))$$

where $q(a)$ is the a quantile of the distribution of super-efficiencies, such that e.g., $q(0.75)$ is the super-efficiency value, below which exist 75% of DSOs .

In addition to these outlier rules, the ordinance prescribes the use of common econometric outlier detection methods like Cook's distance.

The Ordinance also prescribes the *return to scale assumption* to be used in the DEA models of the regulation, namely as a non-decreasing economy of scale, an IRS technology.

The *high level of technical specifications in the German Ordinance* is remarkable and uncommon in an international context. There are several reasons for this. One is probably that it was considered a way to protect the industry against extreme outcomes. The cautious approach of specifying a minimal set of cost drivers and of using the best-of-four approach with an added lower bound of 60% clearly provides some insurance ex-ante to the DSOs about the outcome of future benchmarking analyses. The extensive pre-Ordinance analyses and full scale testing of alternative models and techniques is, of course, also an important pre-requisite. Without such analyses it would not have been possible to design the regulation in such detail nor

to engage in qualified discussion with the industry about alternative approaches. It is worthwhile to note that during the initial analyses leading to the Ordinance, no information was revealed about the efficiency of individual DSOs. Only the general level of efficiency and the distributions of efficiencies were public during this phase.

10.3.4 Model development process

The development of a regulatory benchmarking model is a considerable task due to the diversity of the DSOs involved and the economic consequences that the models may have. Some of the important steps in the German model development are:

Choice of variable standardizations: Choice of accounting standards, cost allocation rules, in/out of scope rules, assets definitions, operating standards etc. are necessary to ensure a good data set from DSOs with different internal practices.

Choice of variable aggregations: Choice of aggregation parameters, like interest and inflation rates, for the calculation of standardized capital costs, and the search for relevant combined cost drivers, using, for example, engineering models, are necessary to reduce the dimensionality of possibly relevant data.

Initial data cleaning: Data collection is an iterative process where definitions are likely to be adjusted and refined and where collected data is constantly monitored by comparing simple KPIs across DSOs and using more advance econometric outlier detection methods.

Average model specification: To complement expert and engineering model results, econometric model specification methods are used to investigate which cost drivers best explain cost and how many cost drivers are necessary.

Frontier model estimations: To determine the relevant DEA and SFA models, they must be estimated, evaluated and tested on full-scale data sets. The starting point is the cost drivers derived from the model specification stage, but the role and significance of these cost drivers must be examined in the frontier models, and alternative specifications derived from using alternative substitutes for the cost drivers must be investigated, taking into account the outlier detecting mechanisms.

Model validation: Extensive second stage analyses are undertaken to see if any of the more than 200 non-included variables should be included. The second stage analyses are typically done using graphical inspection, non-parametric (Kruskal-Wallis) tests for ordinal differences, and truncated regression (Tobit regressions) for cardinal variables. Using the Kruskal-Wallis method, we tested, for example, whether there was an impact on 1) year of cost base, 2) the East-West location of the DSO, and the DSO's possible involvement in water, district heating, gas, or

telecommunication activities. Using Tobit regressions, we tested a series of alternative variables related to cables, connections and meters, substations and transformers, towers, energies delivered, peak flows, decentralized generation, injection points, population changes, soil types, height differences, urbanization, areas etc...

It is worthwhile emphasizing, once again, that model development is not a linear process but rather an iterative one. During the frontier model estimation, for example, one may identify extreme observations that have resulted from data error not captured by the initial data cleaning or the econometric analyses and which may lead to renewed data collection and data corrections. This makes it necessary to redo most steps in an iterative manner.

The non-linear nature of model development constitutes a particular challenge in a regulatory setting where the soundness and details of the process must be documented to allow opposing parties to challenge the regulation in the courtroom.

Also, since corrections of previous steps typically have to repeatedly and since there is also typically a considerable time pressure in the regulatory setting, it is important to organize work appropriately. Scripts to support this can be developed using more advanced software, including R, and are very important and useful for such purposes since they allow massive recalculations in a short period of time and document the calculation steps in great detail.

10.3.5 Model choice

The choice of a benchmarking model in a regulatory context is a multiple criteria problem. There are several objectives, which may conflict with one another. To emphasize this, note at least the following four groups of criteria.

Conceptual: It is important that the model makes conceptual sense both from a theoretical and a practical point of view. The interpretation must be easy and the properties of the model must be natural. This contributes to the acceptance of the model in the industry and provides a safeguard against spurious models developed through data mining and without much understanding of the industry. More precisely, this has to do with the choice of outputs that are natural cost drivers and with functional forms that, for example, have the right return to scale and curvature properties.

Statistical: It is, of course, also important to discipline the search of a good model with classical statistical tests. We seek models that have significant parameters of the right signs and that do not leave a large unexplained variation.

Intuition and experience: Intuition and experience is a less stringent but important safeguard against false model specifications and the over- or underuse of data to draw false conclusions. It is important that the models produce results that are not

that different from the results one would have found in other countries or related industries. Of course, in the usage of such criteria, one also the runs the risk of mistakes. We may screen away extraordinary but true results (Type 1 error) and we may go for a more common set of results based on false models (Type 2 error). The criteria must therefore be used with caution. One aspect of this is that one will tend to be more confident in a specification of inputs and outputs that leads to comparable results in alternative estimation approaches, e.g., in the DEA and SFA model. The experiential basis of this is that when we have a bad model, SFA will see a lot of noise and therefore attribute the deviations from the frontier to noise rather than inefficiency. Efficiencies will therefore be high. DEA, on the other hand, does not distinguish noise and inefficiency, so in a DEA estimation, the companies will look very inefficient. Therefore, results that deviate too drastically in the DEA and SFA estimations may be a sign that the model is not well specified. However, it should be emphasized that the aim is not to generate the same results using a DEA and a SFA estimation. The aim is to find the right model; however, the high correlation between the DEA and SFA results is an indication that the model specification is reasonable. Therefore it also becomes an indirect success criterion.

Regulatory and pragmatic: The regulatory and pragmatic criteria perspective calls for conceptually sound, generally acceptable models as discussed above. Also, the model will ideally be stable in the sense that it does not generate too much fluctuation in the parameters or efficiency evaluations from one year to the next. Otherwise, the regulator will lose credibility and the companies will regard the benchmarking exercise with skepticism. Of course, one will not choose a model simply to make the regulator's life easy, so it is important to remember that similar results are also a sign of a good model specification, cf. the intuitive criteria above. The regulatory perspective also comes into the application of the model. If the model were not good, a high powered incentive scheme, for example, would not be attractive since it would allocate too much risk to the firms. Lastly, let us mention the trivial but very important requirement to comply with the specific conditions laid out in the Ordinance.

Since some of these objectives may conflict we need to make some trade-offs. As an example, it may be that the Ordinance prescribes a cost driver group that in some models is not significant. In that case, there will be a conflict between statistical logic and the law, and we have to make a trade-off in favor of the latter.

Again, the multiple criteria nature of model choice is a particular challenge in regulation. When we have multiple criteria, they may conflict. This means that there is no optimal model that dominates all other models. We have to make trade-offs between different concerns to find a compromise model, to use the language of multiple criteria decision making. Again, such trade-offs can be challenged.

10.3.6 Final model

The final German electricity DSO model used the input and outputs shown in Table 10.2.

Table 10.2 German model of electricity DSOs

Input	Outputs (cost drivers)
Total costs:	yConnections.hs.ms.ns
xTotex or	yCables.circuit.hs.share.cor
xTotex.standard	yLines.circuit.hs.share.cor
	yCables.circuit.ms
	yLines.circuit.ms
	yNet.length.ns
	yPeakload.HSMS.unoccupied.cor
	yPeakload.MSNS.unoccupied.cor
	yArea.supplied.ns
	ySubstations.tot
	yDecentral.prod.cap.tot

From an international perspective, this model specification is comparable in terms of the cost driver coverage included. Regulatory models of electricity DSO generally have cost drivers related to transport work, capacity provision, and service provision. We do not have any transport work cost drivers, but this is in accordance with engineering expectations and is confirmed by both model specification tests and second stage testing. The number of cost drivers is at the high end of what we have used elsewhere.

The DEA models were IRS (NDRS) models as prescribed in the Ordinance, and with the outliers excluded using the two DEA outlier criteria above. In practice, only the last outlier criterion was really effective.

In the SFA models, we used a normed linear specification where the norming constant was yConnections.hs.ms.ns. The reason for norming (deflating) the data was to cope with heteroscedasticity; the absolute excess costs, i.e. the u term, will increase with the size of the company even if the percentage of extra costs are fixed. Likewise, the noise term v is expected to have variance that increases with the size of the DSO. Therefore, the estimated SFA model had the structure

$$\frac{x^k}{y_1^k} = b_1 + b_2\frac{y_2^k}{y_1^k} + \cdots + b_{11}\frac{y_{11}^k}{y_1^k} + u^k + v^k, \qquad k = 1, \ldots, K$$

where u^k is assumed to be truncated normal, v^k to be normal, and where y_1^k is yConnections.hs.ms.ns. Note that the cost, i.e. multiplying through by y_1^k, is assumed to exhibit constant returns to scale.

We could, of course, have handled the heteroscedasticity problem using a log-linear specification, but we did not do so to avoid the specifications curvature prob-

lem ; the output-isoquants in a log-linear specification curve the opposite way than do usual output-isoquants. This is not surprising since the log-linear model corresponds to a Cobb-Douglass model, which is really a production and not a cost function. Besides, conceptually the normed linear model is easy to interpret.

To supplement the analyses, we made sensitivity evaluations of the impact of using a normed linear or a log-linear SFA specification and investigated the impact of using a linear with constant terms which would be more similar to a VRS model. The end results were insensitive to these model variations.

A summary of the resulting efficiency levels are provided in the Table 10.3 below.

Table 10.3 Final efficiencies in German electricity model

Model	Mean	Std.Dev.	Min	$\#E < 0.6$	$\#E = 1$
BestOfTwoTotex	0.898	0.074	0.729	0	40
BestOfTwoTotex.stand.	0.920	0.058	0.795	0	43
BestOfFour	0.922	0.059	0.795	0	49

We see that the resulting efficiency evaluations are high and that with 10 years to catch-up, the yearly requirements are modest. Of course, the catch-up requirements will also be evaluated in terms of the cost elements involved, but there are considerable non-benchmarked cost elements, as we will see below, and a relatively large share of the total costs is Opex.

Although the resulting requirements may seem modest, this is not necessarily a bad outcome for the regulator. First, it may reflect the fact that the German DSOs are relatively efficient, and secondly it may facilitate the institutionalization of model-based regulation. Also, despite the modest estimated average inefficiency of 7.8%, the economic stakes are still considerable at a national level.

Of course, for most companies the stakes are relatively modest and for individual consumers, the stakes are very modest indeed. This actually gives a rationale for central regulation; the individual economic gains are small making it unlikely that individuals will spend many resources challenging the DSO charges.

10.4 DEA based incentive schemes

We will now turn to some more formal integrations of regulation and benchmarking. We first consider DEA based yardstick competition and then DEA based procurement auctions.

The basic problem addressed in this line of research is the following: Given a cross section, a time series or panel information on the multiple inputs and outputs used by K firms

$$(x^k, y^k), \qquad k = 1, \ldots, K$$

what should we ask the firms to do in the future and how should we motivate and compensate them to do so?

The answer to these questions depends on the *organizational context* and in particular on the technological, informational and preferential assumptions of the parties, i.e. the regulator (principal) and firms (agents).

In general, we consider the case where the principal (regulator) faces considerable *uncertainty about the technology*. In a single input multiple output cost setting, the regulator may know that the cost function is increasing and convex, but otherwise have no *a priori* information about the cost structure. In pure moral hazard models, we also assume that the agents face a similar uncertainty.

The general case also empowers agents to take *private actions*, which the principal cannot observe. The action could be to reduce costs or increase the quality of the work done. This leads to a usual *moral hazard problem* since the principal and the agents may conflict as to which actions the agents shall take. The traditional setting depicts the agents as work averse, tempted to rely on their good luck and to explain possibly bad performances with unfavorable circumstances. In general, however, it is simply one way to model the underlying conflicts giving rise to a motivation problem. The conflict might also be that, for example, the medical staff have diverging preferences that induce them to work (too) hard, to treat (groups of) patients below cost and to accommodate requests for multiple treatments .

In some models, we also consider the possibility that the agents have *superior information* about the working conditions before contracting with the principal. A hospital manager may have good information about the primary cost drivers at his hospital while the Ministry of Health may have little information about what causes the total bill to increase. This leads to the classical *adverse selection problem* where an agent will try to extract information rents by claiming to be under less favorable conditions.

Below we report some of the key findings in this literature. For simplicity, we will focus on the single input multiple output case, and interpret the input as a cost and the technology as described by a cost function. This ”cost function” case is the situation that most directly resembles the regulatory problems we have discussed. We note however that similar results are possible for multiple inputs single output production functions as well as for general multiple inputs multiple outputs cases.

10.4.1 Interests and decisions

One of the basic questions is what the firm can decide and how it makes these decisions. This raises a series of issues that are dealt with only superficially in the performance evaluation and incentive literature.

It is common to assume that the principal is risk neutral and that the agent is either risk averse or risk neutral. The principal’s aim is to minimize the costs of inducing the agents to take the desired (hidden) actions in the relevant (hidden) circumstances. An agent’s aim is usually to maximize the utility from payment minus

the dis-utility form private effort. In the combined moral hazard and adverse selection models, we usually make a simplifying assumption about the structure of the agent's trade-offs between effort and payment. We assume that his aim is to maximize a weighted sum of profit and slack:

$$\begin{aligned} U(y^k, B^k) &= u(B^k) - v(y) \\ \text{Firm k's utility} &= \text{Utility from payment} - \text{Cost of effort} \end{aligned}$$

or more specifically,

$$\begin{aligned} U(y^k, B^k) &= (B^k - x^k) - \rho(x^k - c(y^k)) \\ \text{Firm k's utility} &= \text{Profit} + \rho^k \cdot \text{Slack} \end{aligned}$$

where y^k is the outputs produced, B^k is the payment received, and slack is a measure of the extent to which input utilization x^k exceeds the minimal possible $c(y^k)$ and where $1 \geq \rho^k \geq 0$ is the relative value of slack.

We will rely on such assumptions in most of the results below, but we realize that, although widely used in the agency literature, they constitute a stylized caricature of intra-organizational decision making and conflict resolution. This is not satisfactory and is in sharp contrast to the nuanced production description that state-of-the-art performance evaluation techniques like DEA enables. Moreover, recent applications have demonstrated that to derive regulation and incentive schemes with a more sound theoretical basis, we need to know more about what goes on inside the black box of the firm. Only thus can we study, in more detail, the combined use of incentive regulation and regulation by rights and obligations that are used in practice and only in this way can we make valid statements about the speed and path of improvements that a new performance-based scheme may foster. A recent idea of *rational inefficiency* is an attempt to provide a more nuanced view of the preferences involved in the selection of multiple dimensional production plans and slack elements. A discussion of this, however, is beyond the scope of the discussion in this chapter.

10.4.2 Super-efficiency in incentive schemes

One of the first lessons, from the incentive perspective, is that the traditional Farrell score is not useful. The Farrell output efficiency F, for example, gives all units on the relative efficient frontier a score of 1. This severely limits the ability to give high-powered incentives based on Farrell measures. The Farrell measures can give incentives to match others but not to surpass the norm and push out the frontier. Combining this with the multiple dimensional characteristics of the typical DEA model and thereby with the ability to be special in different ways, the *Nash Equilib-*

ria (NE) that can implemented using the Farrell measure will often involve minimal effort and maximal slack.

Figure 10.3 illustrates this. Here, we assume that the cost to the agents is proportional to the length of the production vectors and that payment is decreasing in the F score,

$$F^k > F^{*k} \Rightarrow B^k(F^k) \leq B^k(F^{*k})$$

such that maximal payment is received when a firm is efficient with a score of $F = 1$. If Firm 1 planned to produce at A and moves from A to C, it would get the same payment but use less effort. A is therefore not a best response. Next, Firm 2 could move from the planned B to an easier life in D, again reducing private costs of effort without affecting its payment. This procedure can continue until they both use minimal effort and receive maximal payment.

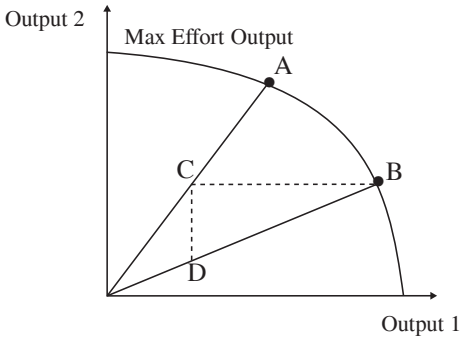


Fig. 10.3 Nash equilibria under Farrell incentives

This somewhat discouraging outcome can easily be remedied by making the payment decreasing in the super-efficiency rather than in the usual output efficiency. In Fig. 10.3, the output-based super efficiency for Firm 1 in A is approximately 0.6, but if the payment is sufficiently decreasing in F^{SUP}, it would not pay to reduce the effort. It does not pay to reduce the effort if the marginal reduction in payment exceeds the marginal decease in the cost of effort.

More generally, *using super-efficiency, one can support the implementation of most plans, even in so-called un-dominated Nash-equilibria*.

10.4.3 Incentives with individual noise

Another fundamental result concerns a pure moral hazard context with ex post evaluations of the performance of the firms when there is

- Considerable technological uncertainty a priori,

- Risk averse firms and
- Individual uncertainty (noise) in the firms' performances.

Technological uncertainty is represented by a large class of *a priori* possible technologies, e.g., the set of production functions that are increasing and concave or the set of functions that are increasing. One can now ask when the DEA frontier is sufficient to write an optimal contract, i.e. when

$$B^{*k} = B^k(x^k, y^k, C^{DEA}(\cdot \mid x^{-k}, y^{-k}))$$

Optimal compensation $= B^k$(Own production, DEA model based on others)

This is the case where optimal relative performance evaluations can be made by comparing the performance of a given firm against the DEA best practice frontier, estimated from the performance of the other firms.

It turns out that i) DEA frontiers support optimal contracts when the distributions of the individual noise terms are exponential or truncated, and that ii) DEA frontiers, based on large samples, support optimal contracts when noise is monotonic, in the sense that small noise terms are more likely than large noise terms. Hence, even when we have individual noise elements and not just the structural uncertainty, which intuitively seems to favor DEA, DEA–based contracts will be optimal for special distributional assumptions and for general assumptions, if the sample is sufficiently large.

10.4.4 Incentives with adverse selection

Another set of results concern combined adverse selection and moral hazard problems with

- Considerable asymmetric information about the technology
- Risk neutral firms,
- Firms seeking to maximize Profit + ρ Slack utility.

The firms are supposed to have superior technological information. In the extreme case, they know the underlying true cost function with certainty, while the regulator only knows the general nature of the cost function. Thus, the regulator may know that there are fixed unit costs of the different outputs but not the exact unit cost because it is the firm's private information. Alternative assumptions may be made about the information available to the regulator. We may assume, for example, that the regulator only knows that the cost function is increasing and convex.

The optimal solution in this case depends on whether the actual costs, i.e. the minimal possible cost plus the slack introduced by the firm, can or cannot be verified and therefore contracted upon.

If the actual costs x cannot be contracted upon, the optimal solution is to use

$$B^{*k}(y^k) = b^k + C^{DEA}(y^k \mid x^{-k}, y^{-k})$$
$$\text{Optimal compensation} = \text{Lump sum} + \text{DEA cost norm ex ante}$$

The size of the lump sum payment depends on the firm's alternatives, i.e. its reservation profit, which in turn depends on profit potentials in other markets or the surplus from contracting with other regulators, for example, private insurance companies. One consequence of this result is that a best way to downsize an organization when there is considerable uncertainty about the cost drivers may be via a lawn-mowing approach where all product types are downsized by the same amount. This situation corresponds to a situation where the only ex ante data is the historical production of the firm in question.

If, instead, we assume that the actual costs of the firm can be contracted upon, the optimal reimbursement scheme becomes

$$B^{*k}(x^k, y^k) = b^k + x^k + \rho^k(C^{DEA}(y^k; x^{-k}, y^{-k}) - x^k)$$
$$\begin{aligned}\text{Optimal compensation} &= \text{Lump sum payment} \\ &+ \text{Actual costs} \\ &+ \rho^k(\text{DEA estimated cost savings})\end{aligned}$$

The structure of this payment scheme can be interpreted as a *DEA based yardstick competition model*: Using the realized performances of the other firms, the regulator creates a cost yardstick against which the regulated firm is evaluated. The regulated firm is allowed to keep a fraction ρ of its saving compared to the yardstick costs as his effective compensation. Fig. 10.4 illustrates this reimbursement scheme.

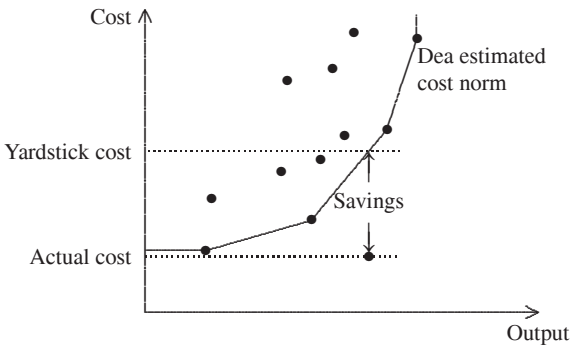


Fig. 10.4 DEA based yardstick competition

These results provide an incentive rationale for using DEA-based revenue cap and yardstick competition systems in contexts where the regulator faces considerable uncertainty about the underlying cost structure. Note that the performance of

the other firms can, in both cases, be interpreted as either historical data, as it is generally used in the revenue cap regulation, or as actual data, as is the idea in the ex post yardstick regulation regime.

10.4.5 Dynamic incentives

In the previous section, we considered incentives for a single period based on historical or current information. Dynamic cases with multiple periods are more complicated since they give rise to new issues like the

- Possibility to accumulate and use new information from one or more firms,
- Need to avoid the Ratchet effect, i.e. deliberate sub-performance in early periods to avoid facing too tough standards in the future and
- Possibility of technical progress (or regress).

The structure of the optimal dynamic scheme is similar to the ones developed above. Thus, the optimal revenue cap for a firm is determined by a DEA-based yardstick norm. Assuming verifiable actual costs and taking into account the generation of new information, the Ratchet effect and the possible technical progress, the optimal scheme becomes

$$B_t^{*k}(x_t^k, y_t^k) = b_t^k + x_t^k + \rho^k(C^{DEA}(y_t^k \mid x_{1-t}^{-k}, y_{1-t}^{-k}) - x_t^k)$$

$$\begin{aligned}\text{Optimal compensation} &= \text{Lump sum payment}\\ &+ \text{ Actual costs}\\ &+ \rho^k(\text{ DEA estimated cost savings})\end{aligned}$$

where $C^{DEA}(y_t^k \mid x_{1-t}^{-k}, y_{1-t}^{-k}) - x_t^k)$ is the DEA-based cost norm that uses all the information from the other firms generated in periods 1 through t. By relying only on information from the other firms in setting the norm, we avoid the Ratchet effect, and by relying on all previous performances, we presume that there is no technical regress.

Of course, the dynamic case can be further extended, e.g. by including incentives to innovate and to share innovative practices. Also, it could be extended to situations where the catch-up capacity is somewhat constrained such that immediate catch-up, as it is assumed here, is avoided.

10.4.6 Bidding incentives

The results summarized above all concern incentives and coordination of activities in view of realized production plans. The realized production plans may be generated ex ante or they may be part of a future multiple agent production context.

An interesting extension of these ideas concerns the possibility of using DEA and related benchmarking techniques to select the winner of a procurement auction and the compensation to provide to the winner. The results above can be extended in this way, although the exercise is non-trivial. The *DEA-based auction* extends the idea of a second price auction to a multiple output case where the services (outputs) offered by the different agents are not the same and where the DEA serves to interpolate a reasonable second price, even in cases where no other bidder is offering the same output profile.

10.4.7 Practical application: DSO regulation in Norway

In 2007 the Norwegian regulator for electricity DSOs, the Norwegian Water Resources and Energy Directorate (NVE), moved from an ex ante revenue cap regulation to a DEA-based yardstick competition regime as it is sketched above with $\rho = 0.6$.

More specifically, the Norwegian revenue cap is determined as

$$R^k(t) = 0.4C^k(t) + 0.6C^k_{DEA}(t-2) + IA^k(t)$$

where R^k is the revenue cap, C^k_{DEA} is the DEA-based cost norm for companies based on data from year $t-2$ and $IA^k(t)$ is the investment addition to take into account the new investments from year t. The actual costs $C^k(t)$ are calculated as

$$C^k(t) = (Opex^k(t-2) + QC^k(t))\frac{CPI(t)}{CPI(t-2)} + pNL^k(t) + DE^k(t-2) + rCap^k(t-2)$$

where QC is quality compensation by firm k to consumers as a consequence of lost load, CPI is the consumer price index, NL is the net-loss, p is the price of power, DE is depreciation, Cap is the capital basis and r is the interest rate on capital set by the regulator.

The cost norm C^k_{DEA} is calculated in two steps. The main calculation is a DEA CRS model with 8 cost drivers covering lines, net stations, delivered energy, numbers of ordinary and vacation users, forests, snow and coastal climate conditions. The second stage is a regression-based second stage correction based on border conditions, decentralized power generation and number of coastal islands in the concession area.

NVE has internationally been a pioneer in the design of model-based regulation of electricity DSOs. In 1991, they introduced Rate of Return Regulation (ROR) and in 1997 they moved to a DEA-based revenue cap regulation that was in place until the introduction of the yardstick regime in 2007. The movement to a yardstick-based regime can be seen as a natural next step in the attempt to mimic a competitive

situation in a natural monopoly industry. Still, the transition from a well-established revenue cap system required careful planning.

One challenge was to convince the industry that a yardstick regime is less risky than an ex ante revenue cap system. The latter enables the companies to predict the future allowed income several years in advance. At first this may seem to be a big advantage but, since it does not include the cost side (except for the use of a more or less arbitrary inflation adjustment), it actually does not protect the companys profit, which should be the main concern for the companies. The yardstick regime offers more insurance because technological progress and costs are estimated directly using the newest possible data.

Another challenge was to calibrate the transition to avoid dramatic changes for any individual firms moving from one benchmarking practice to another.

A third challenge was to enable the firms to close their financial accounts in due time. This is a general challenge of the yardstick competition, and it is a very practical, real challenge. A firm's allowed income for period t can only be calculated after data from all firms have been collected regarding year t. Assuming that the firms are able to deliver this information sometime in the middle of year $t + 1$, the regulator needs at least half a year to validate data and make the calculations. This means that the allowed income for year t will only be known in year $t + 2$. Therefore, in practice, such regulation often works with a time-lag such that the cost norm for period t is based on data from period $t - 2$. This also means that the difference between an ex ante revenue cap and a yardstick-based regime is reduced; the latter becomes similar to a revenue cap with annual updating of the cost norms.

The *structural properties* of the energy industry (firm scale, scope, ownership) may be more important than the details of the regulatory reimbursement schemes. At the same time, the incumbent regulatory regime may have an impact on the structural adjustment, both very directly if the regulators refuse to approve changes in the structure, and indirectly if the payment plans make socially attractive changes non-profitable for the individual firms.

A good example of these problems is the question of how to treat mergers. When payments are correlated with efficiency, the payment plans will tend to discourage mergers in convex models, though they might lead to more outputs being produced with fewer inputs. We have already discussed in Chap. 9 how NVE handles this, by calculating the harmony effect and by compensating a merged firm for the extra requirements corresponding to this effect. At the same time, mergers will tend to affect the performance evaluation basis and may lead to more rents to the firms because the cost norm becomes less demanding by leaving fewer observations in the dataset. The regulator, who considers allowing a merger, must therefore trade-off the gains from improved costs to the firms with the losses from a shrinking information basis. The latter is the regulatory equivalent of the negative market effects in a merger case in a non-regulated sector.

10.5 Summary

Benchmarking can be used to facilitate motivation and contracting. One of the areas where modern benchmarking techniques like DEA and SFA are widely used for motivation purposes is in the regulation of natural monopolies like local or regional electricity and gas distribution systems. In regulatory contexts, the firms generally have superior information about the cost structures, and benchmarking helps the regulator to undermine the firms superior information and, thereby, their ability to extract information rents.

In this chapter, we discussed how different regulations need benchmarking. We saw that price fixation schemes, like a revenue cap system, need benchmarking at least once before every regulatory period, i.e. at least once every 3-5 years, to evaluate the general productivity developments as well as individual incumbent inefficiencies that will determine how much cost reduction the regulator can reasonably request. We also saw that a more advanced regulation like yardstick competition will need yearly benchmarks to evaluate ex post the reasonable costs of the previous year. Lastly, we saw that franchise auctions can make use of benchmarking of the bids to compare different offers across service levels. We also surveyed the systems used in 15 European countries.

As a more specific example, we covered the regulation of German electricity distribution systems operators. We saw how the German approach is cautious. It evaluates every DSO using four different models and relies on the most positive evaluation in setting the allowed income. We also saw how outlier detection based on super-efficiency was part of the regulatory set-up, and we covered the many different steps in a regulatory benchmarking model from the choice of variable standardizations and aggregations, over data cleaning to average model specification, frontier estimations and extensive second stage analyses with the aim of developing a model that is conceptually sound, adheres to general statistical principles, complies with intuition and experience, as well as with regulatory requirements while also taking into account what is feasible and not just desirable. The economic stakes in a regulatory context may be considerable.

Having covered some practical applications, we turned to part of the theoretical basis of DEA-based contracting. We showed that DEA-based contracts may be optimal in some settings, particularly when there is considerable uncertainty about the underlying cost functions. With risk neutral firms, a DEA-based yardstick regime may be the optimal regulation. A specific implementation of this is the new DSO regulation introduced in Norway since 2007.

10.6 Bibliographic notes

Regulation economics was long considered a fairly uninteresting application of industrial organization. Early regulatory theory largely ignored incentive and information issues, drawing heavily on conventional wisdom and industry studies.

This kind of institutional regulatory economics was challenged in the seventies with economists such as Friedman, Baumol, Demsetz and Williamson questioning the organization and succession of natural monopolies. However, the main breakthrough came in the late eighties with the introduction of information economics and agency theory. An authoritative reading in the area is Laffont and Tirole (1993). Littlechild (1983) suggested the price-cap regime, while the idea of yardstick competition goes back to Lazear and Rosen (1981), Nalebuff and Stiglitz (1983) and Shleifer (1985) who show conditions for the implementation of first-best solutions for correlated states of nature. The results carry over, even for imperfectly correlated states of nature Tirole (1988), and as further analyzed using DEA in Bogetoft (1997). Hence, the comparators do not have to be identical, but the relative difference in the exogenous operating conditions has to be known or estimated. Franchise auctions were discussed in, among others, Demsetz (1968) and Laffont and Tirole (1993). The Dutch proposal to let the regulator use benchmarking to put constraints on the acceptable outcomes but to leave the negotiation to industry partners is described in Agrell et al (2007).

Key references to the practical combination of benchmarking and regulation are Agrell and Bogetoft (2001b), Agrell and Bogetoft (2010b) and Coelli et al (2003). A comparison of regulation in the Nordic countries is provided in Agrell et al (2005a)

Relevant references to the German regulation are Agrell and Bogetoft (2007), where we describe the pre-regulation analyses of a series of models to guide the final implementation plan from the regulator as described in Bundesnetzagentur (2007), which was largely transformed into an Ordinance, Government (2007). The 2008 analyses of a new dataset with the aim to serve in the first regulatory period is described in the white paper Agrell and Bogetoft (2008) and the results are summarized in Agrell et al (2008).

The connection between DEA and the formal literature on games was first suggested by Banker (1980) and Banker et al (1989). Linkage with the formal performance evaluation and motivation literature, most notably the agency theory and related regulation and mechanism design literature, has subsequently been the subject of a series of papers including Agrell et al (2002, 2005b), Bogetoft (1994a,b, 1995, 1997, 2000) Bogetoft and Hougaard (2003), Bowlin (1997), Dalen (1996); Dalen and Gomez-Lobo (1997, 2001), Førsund and Kittelsen (1998), Resende (2001), Sheriff (2001), Thanassoulis (2000) and Wunsch (1995). DEA-based auctions were suggested and analyzed in Bogetoft and Nielsen (2008).

The benchmarking model used in the Norwegian yardstick regulation was first developed in Agrell and Bogetoft (2004). The 2010 version of the regulation is summarized in Langset (2009)

Appendix A
Getting Started with R: A Quick Introduction

A.1 Introduction

Throughout this book, we use R to perform our calculations and applications. R is a free software environment for data analysis and graphics. R is a scripting language that provides a degree of control that a menu-based system cannot readily provide. In R, a user can easily apply an output from a function or method as an input for a subsequent function or method, and, therefore, it is always possible to perform further calculations with the results of a benchmark function or statistical analysis.

In R, it is easy to combine existing methods and write new methods that can simplify an analysis and can be shared with other users. These features have made it possible for users to write functions that solve all of the benchmarking problems in this book. These functions were then added to a documented add-on package called Benchmarking, which is freely available for others to use.

This chapter is not an introduction to all of R, but to the selected parts of R that we use in the book. For further introductions, see *An Introduction to R*, which is available on `http://www.r-project.org` under Documentation/Manuals. Moreover, many introductions to R also serve as introductions to statistics, for example Dalgaard (2002). The *A Beginners Guide to R* (Zuur et al, 2009) is a detailed introduction to the everyday use of R that contains more than enough information for a new user to successfully implement the methods in this book.

A.2 Getting and installing R

Download the newest version of R from `www.r-project.org`. Manuals and other materials are also available on this website.

1. Open the page `www.r-project.org` and select the link `CRAN` on the left of the page, below `Download, Packages`.
2. In the list of CRAN mirrors click on the one nearest you.

P. Bogetoft and L. Otto, *Benchmarking with DEA, SFA, and R*, International Series in Operations Research & Management Science 157, DOI 10.1007/978-1-4419-7961-2,

3. Click on your operating system under Download and Install R.
4. The next step depends on the selected operating system. In any case, choose and download the latest binary version, which is almost 40 MB.
   Windows users should choose base and then `Download R 2.11.1 for Windows`, which is the version number at the time of writing. It is also possible to choose the version `r-patched snapshot build`, which includes the latest bug fixes, patches.
5. Download and run the file to install R. This step depends on each operating system's standards.

Once R is installed, it can be used.

A.3 An introductory R session

Start R, and a console windows in which you can write individual commands on the lines. R uses ">" to prompt for an input, a new command and "+" to prompt for a continued line with input. When a user enters a number into R, R writes the number back on the console. When a user writes an expression, R calculates the result and writes in on the console. It is also possible to assign values to variables with the assignment operator.,"<-":

```
> 3
[1] 3
> 3 + 5
[1] 8
> 2 * (3 + 5)
[1] 16
> a <- 3
> A <- 10
> b <- 5
> a + b
[1] 8
> A + b
[1] 15
```

It should be noted that the variables `a` and `A` are different, that is, R distinguishes between lower and upper cases in all kinds of variables. Variables can also be vectors or arrays. They can also be constructed by concatenation with the operator `c` and used in calculations:

```
> x <- c(1, 3, 6, 8)
[1] 1 3 6 8
> y <- c(99, 3, 67, 103)
[1]  99   3  67 103
> x + y
[1] 100   6  73 111
> x + 10
[1] 11 13 16 18
```

```
> xx <- c(x, x)
[1] 1 3 6 8 1 3 6 8
> xx + y
[1] 100   6  73 111 100   6  73 111
> z <- c(1, 2, 3)
[1] 1 2 3
> z + y
[1] 100   5  70 104
Warning message:
In z + y : longer object length is not a multiple of
           shorter object length
```

Users can add arrays and can also add scalars to arrays. When a user adds arrays of different lengths, R recycles the shortest, but, as the above example shows, it provides a warning if the longer array is not a multiple of the shorter.

There is no end to the possible uses for concatenation and assignment.

```
> z1 <- c(x, 0, x)
[1] 1 3 6 8 0 1 3 6 8
> z2 <- c(x, y)
[1]   1   3   6   8  99   3  67 103
```

The result of concatenation is a long, one-dimensional array.

Data can be organized into a named list of variables, which is called a data frame in R. The function to create a data frame is `data.frame`.

```
>   xyDat <- data.frame(x,y, digits=c(1,2,3,4),
+         numb=c("one","two","three","four"))
>   xyDat
  x   y digits  numb
1 1  99      1   one
2 3   3      2   two
3 6  67      3 three
4 8 103      4  four
```

It should be noted that the variables (i.e., the columns in the data frame) are given (new) names with = in the `data.frame`.

R can read data from a file and save them into a data frame. For instance, the data in the file `smallData.txt` with the following contents

```
labor output industry
100      75    manufact
200     100    manufact
300     300    service
500     400    agricult
```

can be stored in the data frame `d` with the R commands

```
> d <- read.table("smallData.txt", header = TRUE)
> d
  labor output industry
1   100     75 manufact
2   200    100 manufact
```

```
3   300    300  service
4   500    400 agricult
> names(d)
[1] "labor"    "output"   "industry"
> d$labor
[1] 100 200 300 500
```

where `header=TRUE` says that the first line contains the variable names. Furthermore, `d$labor` refers to the variable `labor`, which is the component `labor`, in the data frame `d`. The subsection A.5 presents several other methods for reading data.

As can be seen from above, data are not limited to numerical data; they can include categorical data, such as strings of names or indications of data subcategories. In R, such variables are called factors and are often useful for comparing or defining groups in a benchmark analysis. For the data above, we can treat `industry` as a factor. This factor, `industry`, has three levels - the three different factor values - which are confirmed when the variable is printed and when the function `levels` retrieves the levels.

```
> d$industry
[1] manufact manufact service  agricult
Levels: agricult manufact service
> levels(d$industry)
[1] "agricult" "manufact" "service"
```

An index can be used to reference a specific part of a dataset, array, or matrix. Indices are given in brackets. For example,

```
> d[3, 2]
[1] 300
> d[2, ]
  labor output industry
2   200    100 manufact
> d$labor[2]
[1] 200
```

If one index is left out, the whole row or column is returned.

Matrix calculations are often useful in benchmarking. Numerical data can be transferred into a matrix, to which all the known matrix functions apply. Categorical values (i.e., factors), like string values, can also be part of a matrix, but in this case the numerical matrix methods do not apply.

Several arrays of the same length can be turned into a matrix, as can a long array:

```
> x <- 1:6
[1] 1 2 3 4 5 6
> X <- matrix(x, ncol = 2)
     [,1] [,2]
[1,]    1    4
[2,]    2    5
[3,]    3    6
> Y <- matrix(c(11:16, 21, 22, 29), nrow = 3)
```

```
     [,1] [,2] [,3]
[1,]   11   14   21
[2,]   12   15   22
[3,]   13   16   29
> Z <- matrix(c(11:16, 21, 22, 23), ncol = 3, byrow = T)
     [,1] [,2] [,3]
[1,]   11   12   13
[2,]   14   15   16
[3,]   21   22   23
```

It should be noted that `1:6` is shorthand for all the numbers from the first to the last value (i.e., from 1 to 6). The option `byrow=T` denotes that the data for the matrix should be filled by rows.

Matrices can also be created by combining columns and rows, respectively. For example,

```
> cbind(d$output, d$labor)
     [,1] [,2]
[1,]   75  100
[2,]  100  200
[3,]  300  300
[4,]  400  500
> rbind(output = d$output, newName = d$labor)
        [,1] [,2] [,3] [,4]
output    75  100  300  400
newName  100  200  300  500
```

The results can be assigned to variables by the assignment operator "<-", which can also give the columns and rows names.

The inverse of a matrix can be found as

```
> solve(Y)
           [,1]       [,2]       [,3]
[1,] -4.6111111  3.8888889  0.3888889
[2,]  3.4444444 -2.5555556 -0.5555556
[3,]  0.1666667 -0.3333333  0.1666667
```

For data in matrices, all the standard matrix operations are available, for example, `%*%` for matrix multiplication. It should be noted that if `A` and `B` are matrices, then `A*B` is an element according to element multiplication and `A%*%B` is a matrix multiplication (i.e., the inner product of matrices), as can be seen from the following

```
> Y * solve(Y)
           [,1]       [,2]       [,3]
[1,] -50.722222  54.444444   8.166667
[2,]  41.333333 -38.333333 -12.222222
[3,]   2.166667  -5.333333   4.833333
> Y %*% solve(Y)
             [,1]          [,2]          [,3]
[1,] 1.000000e+00 -9.381385e-15 -1.082467e-15
[2,] 8.382184e-15  1.000000e+00 -5.551115e-17
[3,] 4.468648e-15 -1.831868e-15  1.000000e+00
```

where the non-zero values of the diagonal elements in the last matrix are due to round-off errors in the calculations in R, which is one example of R's precision.

Many other functions for matrices exist in R, but it is important to note that R often offers a better, faster and more efficient method or function to handle a specific problem than the calculations in a series of matrix operations do. For example, R does not use `solve(t(X)%*%X)%*%t(X)%*%y` to obtain the parameters for a linear regression; instead, it uses the method `lm(y~X)` for linear regression models. Most of these specific methods have a functionality to obtain frequently used, derived variables.

The result of an evaluated expression, whether a simple expression or the result of a function or method, is either printed on the screen (output file) or assigned to a variable, as in the following example.

```
> matrix(1:6, nrow = 2)
     [,1] [,2] [,3]
[1,]    1    3    5
[2,]    2    4    6
> x <- matrix(1:6, nrow = 2)
> x
     [,1] [,2] [,3]
[1,]    1    3    5
[2,]    2    4    6
```

The result of executing the variable is a print-out of the variable's contents. If the variable is the result of a function, like `dea` or `sfa`, then the variable or object might contain more information than what is printed. The user may access the other parts of the object may be accessed with a method or a component.

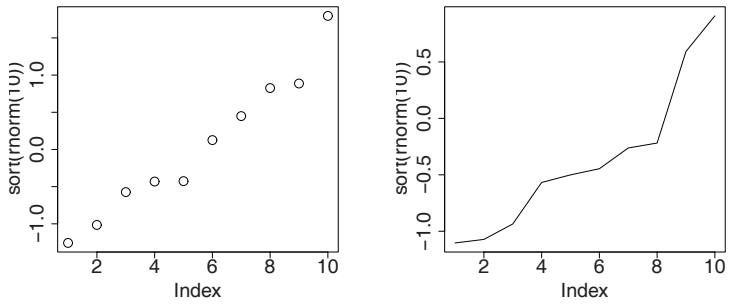
The function `dea` in the add-on package Benchmarking produces a list of variables (for more on packages, see Sect. A.3.1). In the following commands, we show the first four variables (names) in the object `e` and access the component `eff` both with a function and direct access.

```
> library(Benchmarking)
> x <- matrix(c(10, 20, 30, 50), ncol = 1)
> y <- matrix(c(7, 10, 30, 40), ncol = 1)
> e <- dea(x, y)
> names(e)[1:4]
[1] "eff"    "lambda" "objval" "RTS"
> eff(e)
             E
[1,] 1.0000000
[2,] 0.6304348
[3,] 1.0000000
[4,] 1.0000000
> e$eff
[1] 1.0000000 0.6304348 1.0000000 1.0000000
```

Many functions in R produce structures or lists defined by components, and the online-help gives information on accessing the different parts.

R performs impressively on graphics. For example:

```
> plot(sort(rnorm(10)))
> plot(sort(rnorm(10)), type = "l")
```



. The function `rnorm(10)` assigns 10 random, standard, normally distributed values; the function `plot` makes a plot; and the `sort` sorts the values.

The default plotting method is with points, but the option `type="l"` creates a line instead. Most introductions to R contain a description of the plotting methods and explain how to annotate a plot. Adding points or lines to a plot uses `points`, `lines`, and `abline`. There are many specialized plot functions. We use `dea.plot` to plot technology sets, which can almost always be combined with the standard plot options for line type, plotting character and size, and annotations (see the documentation for `dea.plot` in the manual for package Bechmarking).

Online help and documentation for a method, function, or package can be found with either `help("name")` or `?name`. These methods only provide documentation for loaded packages. For unspecified information, the `??efficiency` searches for anything with efficiency installed but not necessarily in loaded packages.

To quit the R program, use the command

```
> q()
```

and select “no” in response to the prompt to save the workspace. It is of course also possible to exit from the menu in the graphical user interface.

A.3.1 Packages

Not everything can be done in basic R, but, with add-on packages, almost everything can. Users can find and download new packages from CRAN, which can be accessed via the link on R’s homepage. For most users, it is probably easier to install new packages in an R session, while connected to the Internet, using a submenu under the Package menu. Installed packages can be kept up to date with the function `update.packages()`, or, perhaps more easily, with the use of a submenu under the Package menu. Of course, this menu can only be used when the computer is connected to the Internet.

Once a package is installed, it is saved to the computer's hard-disk. However, a package also has to be loaded to be used in a given R session, which can be performed with the method `library`. For example:

```
> library(Benchmarking)
```

to load the installed package `Benchmarking`. In Windows and OS X, this task can also be performed from the menu Packages, Load packages .... A user should load the packages he wishes to use for each new R session.

A.3.2 Scripts

Scripts are R commands that are collected in a file so that the user can run them at a later date without having to re-type the commands into the console window. A script file is a pure text file, without any formation, that can be opened, edited, and saved from within R. It is a good idea for users to take advantage of scripts, especially those who do not plan to use R every day; they should keep old scripts that can be used as a starting point for new scripts on new projects.

A.3.3 Files in R

Several kind of files that are relevant to R; here, we list those related to commands and data, which are available under the File menu.

Script Opens a window with a file whose contents can be executed by the menu Edit, Run line or by selection with a right mouse click or Ctrl+R in OS X. Changes to a script must be saved via the menu File, Save or by Ctrl+S.

Workspace Saves variables and functions in a binary file that only R can read. The commands themselves are not saved.

History Saves all the executed commands, even the faulty ones, but not the corresponding output. It can be edited by any text editor, saved as a text-file, and then opened as a script file. This file offers a way for users to make new scripts after having found the right commands.

Using Word to edit R files is not recommended, but if you do be sure to save the file a pure text file without any Word formatting, which R treats as errors.

The graphical window can be saved with a right click on the mouse or, when the graphical window is the chosen window, via the menu File and Save As. Of the many available graphical formats, encapsulated PostScript is suitable for LaTeX, and png is suitable for Word.

A.4 Changing the appearance of graphs

As mentioned, the function `plot` plots points by default. To plot lines, add the option `type=''l''`. To add further lines or points to an existing graph, use the function `lines` or `points`. The function `abline` adds one or more straight lines to the current plot.

The following options are available for the `plot` and the other graph functions:

lty Line type, either integer or string: 1: solid (default), 2: dashed, 3: dotted, 4: dotdash, 5: longdash, 6: twodash; use `lty=2` or `lty="dashed"`.
lwd Width of line, default is 1.0.
pch Plotting symbol, a number from 1 to 25; use `pch=16`
cex The relative size of text and plotting symbols, the default is 1.0.
main The title of the plot, `main="Heading"`
xlab Label for the x axis, `xlab="Label for x axis"`
ylab Label for the y axis, `ylab="Label for y axis"`
col Specify plotting color (see documentation or the function `colors()`)

R offers many other possibilities to control graphing and graph appearance (see the documentation for `plot` and `par`).

A.5 Reading data into R

Users can transfer data into R using various formats. In Windows, the package xlsReadWrite can be used to read Excel files directly; the method is `read.xls`. An alternative for all operating systems when data is a csv file or can be saved as one via Excel is the method `read.csv`,. Of course, the method `read.table` can read ordinary numbers and text in a matrix or table in an ascii file.

The data are read and put into a data frame `d` with the commands

```
d <- read.table("filename")
d <- read.csv("filename")
```

For all the reading methods, there is an option `header`—`colNames` in `read.xls`, which indicates whether the file contains the names of the variables in its first line. When the first line lists the variable names, `header=TRUE` should be performed; users do not have to do so for `read.csv` because it is the default. The methods offer a series of other options that users can peruse by executing the command `?read.csv`, `?read.table`, or `?read.xls`.

A.5.1 Reading data from Excel

Data from an Excel file can be read into R using either the clipboard, a special function for xls files, or a csv file.

To use the clipboard, Excel must be open. Choose the data and make a copy; the data can then can be read into R by

```
read.table("clipboard",header=TRUE)
```

.

The package `xlsReadWrite` can read directly from an Excel file, even if Excel is not open:

```
library(xlsReadWrite)
dat <- read.xls("filename.xls",sheet="data")
```

The first line in the Excel file should be a row of names; if it is not, the option `colNames=FALSE` should be used.

R can also read data from a csv file. In Excel, save the file as a csv file and read it into R with the command `read.csv`.

A.6 Benchmarking methods

Table A.1 lists many of the methods used in this book that are available in the package Benchmarking. The table only lists the main options for the methods; the remaining options and their descriptions are available in the R documentation for this method.

A.7 A first R script for benchmarking

The following section presents an example of a small script for performing a benchmark analysis with a small data set. The main chapters discuss all the functions in detail, especially Chapts. 4 and 5.

```
library(Benchmarking)
# Get the data, normally read from a file
x <- matrix(c(100,200,300,500,100,600),ncol=1)
y <- matrix(c( 75,100,300,400, 50,400),ncol=1)

# Plot of different technologies with a title
dea.plot(x,y, txt=1:dim(x)[1],main="Basic␣plot␣of␣frontier")
dea.plot(x,y,RTS="crs",add=TRUE,lty="dashed")
```

Table A.1 Methods and main options in Benchmarking

Method	Description
dea	Calculates Farrell efficiency; the most important options are shown below
sdea	Super-efficiency, uses dea above
dea.plot	General plot of technology for 2 goods
dea.plot.isoquant	Plot technology for 2 inputs, isoquant—implicitly assumes that outputs are the same for all firms
dea.plot.frontier	Plots technology for 1 input and 1 output
dea.dual	Calculates efficiency and handles dual restrictions
make.merge	Calculates aggregation matrices and aggregate input and output matrices
dea.merge	Calculates efficiencies related to merger analysis
dea.boot	Bootstrap of DEA models, not available; use `boot.sw98` from the package FEAR
sfa	SFA, input and output as matrices

Main options with defaults	Description
X	Input where efficiency is to be calculated; a $K \times m$ matrix of observations of K firms with m inputs
Y	Output where efficiency is to be calculated; a $K \times n$ matrix of observations of K firms with n outputs
RTS="vrs"	Returns to scale assumption, fdh, vrs, drs, crs, irs, add
ORIENTATION="in"	Input, output og graph oriented efficiency; 'in', 'out', or 'graph'

Note: The options depends on the method; the main options for most of the methods are given above. Each method has further options for its specialization; for details, see the online documentation.

```
# Calculate efficiency,
dea(x,y, RTS="crs", ORIENTATION="in")
e <- dea(x,y, RTS="vrs", ORIENTATION="in")
e
# Show the peers
peers(e)
# Show the weights for the peers
lambda(e)

# Calculate how much input could be saved if all firms were
```

```
# efficient, per firm
sav <- (1-e$eff)*x
sav
# and for all firms aggregated
sum(sav)
# The percentage that could be saved
100 * sum(sav)/sum(x)

# Taking care of slacks
# Phase one: ordinary dea
e <- dea(x,y)
# Phase two: calculate slacks
sl <- slack(x,y,e)
data.frame(eff(e),eff(sl),sl$slack,sl$sx,sl$sy,lambda(sl))
peers(e)
peers(sl)

# The two phases in one function call
e2 <- dea(x,y,SLACK=TRUE)
print(e2)
data.frame(eff(e2),e2$slack,e2$sx,e2$sy,lambda(e2))
peers(e2)
```

A.8 Other packages for benchmarking in R

The following table lists of some of the relevant methods in packages for efficiency analysis that are available in R. This section only presents methods for handling the problems that are mentioned in the main text; the packages include more methods than are mentioned here.

FEAR

The FEAR package by Wilson (2008) is fast and easy to use. The interface that is used for methods in Benchmarking is, approximately, an expanded version of the FEAR package's interface. The package is not freely available in R but can be acquired from the author via the website, `http://www.clemson.edu/economics/faculty/wilson/Software/FEAR/fear.html`. It is only available for Windows and Linux. The efficiencies in FEAR are Shephard distance functions, not the Farrell efficiencies that we use most frequently in this book. However, Farrell efficiencies can be easily generated via `1/FEAR::dea(X,Y)`. FEAR does not provide peers for units or dual values. The source code is not available, and therefore it is not possible to make corrections for missing methods. Input is in

Table A.2 Other packages and methods

FEAR	*Frontier Efficiency Analysis with R*
dea	Compute Shephard DEA efficiency estimates; only VRS, DRS, and CRS
fdh	Compute Shephard FDH efficiency estimates
boot.sw98	Homogeneous bootstrap for Shephard distance Functions
ap	Outlier detection for non-parametric frontier models
ap.plot	Produce log-ratio plot for outlier analysis
malmquist	Malmquist productivity indices
frontier	*Stochastic Frontier Analysis based on Tim Coelli's program Frontier 4.1*
sfa	Maximum likelihood estimation of stochastic frontier production and cost functions

the dimension firm $\times$ good, which is transposed compared to the ordinary use in R. Therefore, it the data are in a data frame, then the input and output matrices must be transposed after they are selected from the data frame.

To use FEAR and Benchmarking in the same R session, it is necessary to use the packagename plus "::" as a prefix for the function `dea`. If FEAR is loaded first, the `FEAR::dea` should be used; if Benchmarking is loaded first, the `Benchmarking::dea` should be used. The other functions are not influenced.

frontier

This package is based on Tim Coelli's program Frontier 4.1, which was written in Fortran. Unfortunately, the program seems to use a substandard optimization routine by todays standard because it occasionally fails or provides the wrong parameters for problems that better optimization routines would deal with successfully. The parameterization is different from the method that we use, and therefore some of the parameters and the calculations of variance components differ. A good quality interface corresponding to the interface for the linear models, lm and glm, is available and makes it possible to specify a linear component for the efficiency part. The methods can handle panel data.

To use frontier and Benchmarking in the same R session, the same "::" procedure should be implemented for the function `sfa` in the frontier package, as mentioned above for the package FEAR.

A.9 Bibliographic notes

Further documentation on R can be found in the manual section on the R homepage http://www.r-project.org. *An Introduction to R* is a good way to start; we advise skimming through the first section: Introduction and Preliminaries. Several books, such as *A Beginner's Guide to R (Use R)* (Zuur et al, 2009) deal with almost the same subjects as the *Introduction to R* but in more details. Furthermore, several books offer an introduction to R and statistics, such as Dalgaard, *Introductory Statistics with R*, 2nd edition 2008 Dalgaard (2002) and the more advanced Venables and Ripley (1999), which is available in many editions. Further information on data manipulations, reading, databases, dates, factors, and reshaping data is available in Spector (2008). Other books introduce specialized areas of R, such as data manipulation, non-linear regression, and econometrics.

A complete survey of graphics in R is available in Murrell (2006), but it is not an introduction to this subject.

References

Afriat SN (1972) Efficiency estimation of production functions. International Economic Review 13(568–598)

Agrell P, Bogetoft P (2007) Development of benchmarking models for german electricity and gas distribution. Consultation report, Bundesnetzagentur, Bonn, Germany

Agrell P, Bogetoft P (2009) International benchmarking of electricity transmission system operators - e3grid project. Consultation report, open version, Council of European Energy Regulators

Agrell P, Bogetoft P, RHalbersma, MCMikkers (2007) Yardstick competition for multi-product hospitals. NZa Research Paper 2007/1, NZa, Netherlands

Agrell PJ, Bogetoft P (2000) Ekonomisk nätbesiktning. final report stem. Tech. rep., SUMICSID AB. (In Swedish)

Agrell PJ, Bogetoft P (2001a) Incentive regulation. Working Paper

Agrell PJ, Bogetoft P (2001b) Should health regulators use DEA? In: Fidalgo Eea (ed) Coordinacion e Incentivos en Sanidad, Asociasion de Economia de la Salud, Barcelona, pp133-154

Agrell PJ, Bogetoft P (2003) Norm models. Consultation report, Norwegian Water Resources and Energy Directorate (NVE)

Agrell PJ, Bogetoft P (2004) Nve network cost efficiency model,. Tech. rep., Norwegian Energy Directorate NVE

Agrell PJ, Bogetoft P (2008) Electricity and gas dso benchmarking whitepaper. Consulattion report, Bundesnetzagentur

Agrell PJ, Bogetoft P (2010a) Benchmarking of german gas transmission system operators. Consultation report, Bundesnetzagentur (BNetzA)

Agrell PJ, Bogetoft P (2010b) A primer on regulation and benchmarking with examples from network industries. Tech. Rep. version 05, SUMICSID AB

Agrell PJ, Tind J (2001) A dual approach to noconvex frontier models. Journal of Productivity Analysis 16:129–147

Agrell PJ, Bogetoft P, Tind J (2002) Incentive plans for productive efficiency, innovation and learning. International Journal of Production Economics 78:1–11

Agrell PJ, Bogetoft P, Bjørndalen J, Vanhanen J, Syrjänen M (2005a) Nemesys subproject a: System analysis. Consultation report, Nordenergi

Agrell PJ, Bogetoft P, Tind J (2005b) Dea and dynamic yardstick competition in scandinavian electricity distribution. Journal of Productivity Analysis 23:173–201

Agrell PJ, Bogetoft P, Cullmann A, von Hirschhausen C, Neumann A, Walter M (2008) Ergebnisdokumentation: Bestimmung der effizienzwerte verteilernetzbetreiber strom. Consultation report, Bundesnetzagentur

Aigner DJ, Chu SF (1968) On Estimating the Industry Production Function. American Economic Review 58:826–839

Aigner DJ, Lovell CAK, Schmidt P (1977) Formulation and Estimation of Stochastic Frontier Production Function Models. Journal of Econometrics 6:21–37
Andersen J, Bogetoft P (2007) Gains from Quota Trade: Theoretical models and an application to the Danish Fishery. European Review of Agricultural Economics 34(1):105–127
Andersen P, Petersen NC (1993) A procedure for ranking efficient units in data envelopment analysis. Management Science 39(10):1261–1264
Andrews DF, Pregibon D (1978) Finding the outliers that matter. Journal of the Royal Statistical Society, Series B 40(1):85–93
Asmild M, Hougaard JL, Kronborg D, Kvist HK (2003) Measuring inefficiency via potential improvements. Journal of Productivity Analysis 19:59–76
Asmild M, Bogetoft P, Hougaard JL (2008) Rationalising inefficiency: A study of canadian bank branches. Working Paper, FOI, Copenhagen University
Atkinson AC (1985) Plots, Transformations and Regression. Clarendon Press: Oxford
Banker RD (1980) A game theoretic approach to measuring efficiency. European Journal of Operational Research 5:262–268
Banker RD (1984) Estimating most productive scale size using data envelopment analysis. European Journal of Operational Research 17(1):35–54
Banker RD (1993) Maximum likelihood, consistency and data envelopment analysis: A statistical foundation. Management Science 39(10):1265–1273
Banker RD (1996) Hypothesis tests using data envelopment analysis. Journal of Productivity Analysis 7:139–159
Banker RD, Chang H (2006) The super-efficiency procedure for outlier identification, not for ranking efficient units. European Journal of Operational Research 175(2):1311–1320
Banker RD, Morey RC (1986) Efficiency analysis for exogenously fixed inputs and outputs. Operations Research 34(4):513–521
Banker RD, Thrall R (1992) Estimation of returns to scale using data envelopment analysis. European Journal of Operational Research 62:74–84
Banker RD, Charnes A, Cooper WW (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis. Management Science 30:1078–1092
Banker RD, Charnes A, Cooper WW, Clarke R (1989) Constrained game formulations and interpretations for data envelopment analysis. European Journal of Operational Research 40:299–308
Battese G, Coelli T (1992) Frontier production functions, technical efficiency and panel data: With application to paddy farmers in india. Journal of Productivity Analysis 3:153–169
Battese GE, Coelli TJ (1988) Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. Journal of Econometrics 38(3):387–399
Belsley DA, Kuh E, Welsch RE (1980) Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. Wiley: New York
Bogetoft P (1990) Strategic responses to dea-control - a game theoretical analysis. Tech. rep., Copenhagen Business School
Bogetoft P (1994a) Incentive Efficient Production Frontiers: An Agency Perspective on DEA. Management Science 40:959–968
Bogetoft P (1994b) Non-Cooperative Planning Theory. Springer-Verlag
Bogetoft P (1995) Incentives and Productivity Measurements. International Journal of Production Economics 39:67–81
Bogetoft P (1996) DEA on Relaxed Convexity Assumptions. Management Science 42:457–465
Bogetoft P (1997) DEA-based yardstick competition: The optimality of best practice regulation. Annals of Operations Research 73:277–298
Bogetoft P (2000) DEA and activity planning under asymmetric information. Journal of Productivity Analysis 13:7–48
Bogetoft P, Gammeltvedt TE (2006) Mergers in norwegian electricity distribution: A cost saving exercise? Working paper, NVE, Norway
Bogetoft P, Hougaard JL (1999) Efficiency evaluation based on potential (non-proportional) improvements. Journal of Productivity Analysis 12:233–247

Bogetoft P, Hougaard JL (2003) Rational inefficiencies. Journal of Productivity Analysis 20:243–271
Bogetoft P, Hougaard JL (2004) Super efficiency evaluation based on potential slack. European Journal of Operational Research 152:14–21
Bogetoft P, Katona K (2008) Efficiency gains from mergers in the healthcare sector. Tech. rep., Nederlandse Zorgautoriteit NZA
Bogetoft P, Nielsen K (2004) Monitoring farm, herd and cow performance - efficiency analyses. Tech. rep., Royal Agricultural University and www.kvægforskning.dk
Bogetoft P, Nielsen K (2005) Internet Based Benchmarking. Journal of Group Decision and Negotiation 14(3):195–215
Bogetoft P, Nielsen K (2008) DEA based auctions. European Journal of Operational Research 184:685–700
Bogetoft P, Otto L (2010) Benchmark package. Tech. rep., R
Bogetoft P, Pruzan P (1991) Planning with Multiple Criteria, 1st edn. North-Holland
Bogetoft P, Wang D (2005) Estimating the Potential Gains from Mergers. Journal of Productivity Analysis 23:145–171
Bogetoft P, Tama J, Tind J (2000) Convex input and output projections of nonconvex production possibility sets. Management Science 46:858–869
Bogetoft P, Strange N, Thorsen BJ (2003) Efficiency and Merger Gains in The Danish Forestry Extension Service. Forest Science 49(4):585–595
Bogetoft P, Fried H, Eeckaut PV (2004) Power benchmarking: What's wrong with traditional benchmarking and how to do it right. Tech. rep., Credit Union Research and Advice, Credit Union National Association, http://thepoint.cuna.org/
Bogetoft P, Bramsen JM, Nielsen K (2006a) Balanced Benchmarking. International Journal of Business Performance Management 8(4):274–289
Bogetoft P, Färe R, Obel B (2006b) Allocative Efficiency of Technically Inefficient Production Units. European Journal of Operational Research 168(2):450–462
Bogetoft P, Boye K, Neergaard-Petersen H, Nielsen K (2007a) Reallocating sugar beet contracts: can sugar production survive in Denmark. European Review of Agricultural Economics 34(1):1–20
Bogetoft P, Fried HO, Eeckaut PV (2007b) The university benchmarker: An interactive computer approach. In: Bonaccorsi A, Daraio C (eds) Universities And Strategic Knowledge Creation, Edward Elgar Publishing, chap 14
Bogetoft P, Christensen D, Damgård I, Geisler M, Jakobsen T, Krøigaard M, Nielsen J, Nielsen J, Nielsen K, Pagter J, et al (2009) Secure multiparty computation goes live. Financial Cryptography and Data Security pp 325–343
Bogetoft P, Kristensen T, Pedersen KM (2010) Potential gains from hospital mergers in denmark. Health Care Management Science To Appear
Bowlin W (1997) A proposal for designing employment contracts for government managers. Socio-Economic Planning Sciences 31:205–216
Brännlund R, Färe R, Grosskopf S (1995) Environmental regulation and profitability: An application to swedish pulp and paper mills. Environmental and Resource Economics 6
Brännlund R, Chung Y, Färe R, Grosskopf S (1998) Emissions Trading and Profitability: The Swedish Pulp and Paper Industry. Environmental and Resource Economics 12:345–356
Bundesnetzagentur (2007) Bericht der bundesnetzagentur nach § 112a enwg zur einführung der anreizregulierung nach § 21a enwg. Report, Bundesnetzagentur
Caves DW, Christensen LR, Diewert WE (1982) The economic theory of index numbers and the measurement of input, output, and productivity. Econometrica 50(6):1393–1414
Chambers JM, Hastie TJ (eds) (1992) Statistical Models in S. Wadsworth & Brooks
Chambers RG (1988) Applied production analysis: A dual approach. Cambridge: Cambridge University Press
Chambers RG, Chung Y, Fare R (1996) Benefit and distance functions. Journal of Economic Theory 70(2):407–419

Chambers RG, Chung Y, Färe R (1998) Profit, directional distance functions, and nerlovian efficiency. Journal of Optimization Theory and Application 2:351–364
Chang K, Guh Y (1991) Linear production functions and the data envelopment analysis. European Journal of Operational Research 52:215–233
Chang KP (1999) Measuring efficiency with quasiconcave production frontiers. European Journal of Operational Research 115:497–506
Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision making units. European Journal of Operational Research 2:429–444
Charnes A, Cooper WW, Rhodes E (1979) Short Communication: Measuring the Efficiency of Decision Making Units. European Journal of Operational Research 3:339
Charnes A, Cooper WW, 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, Cooper WW, Wei QL, Huang ZM (1989) Cone ratio data envelopment analysis and multi-objective programming. International Journal of Systems Science 20:1099–1118
Charnes A, Cooper WW, Lewin AY, Seiford LM (1995) Data Envelopment Analysis: Theory, Methodology and Applications. Kluwer Academic Publishers, Boston, USA
Christensen LR, Jorgenson DW, Lau LJ (1973) Transcendental logarithmic production frontiers. Review of Economics and Statistics 55:28–45
Coelli T, Prasada Rao DS, Battese G (1998a) An Introduction to Efficiency and Productivity Analysis. Kluwer Academic Publishers
Coelli T, Rao DP, Battese GE (1998b) An Introduction to Efficency and Productivity Analysis. Kluwer, Boston, USA
Coelli T, Estache A, Perelman S, Trujillo L (2003) A primer on efficiency measurement for utilities and transport regulators. Tech. Rep. 129, World Bank Publications
Cooper WW, Seiford LM, Tone K (2007) Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software, 2nd edn. Springer, Secaucus, NJ, USA
Cox D, Hinkley D (1974) Theoretical Statistics. Chapman and Hall, London
Credit Union National Association (2010) Cub. URL http://advice.cuna.org/cu_benchmarker.html
Dalen DM (1996) Strategic responses to relative evaluation of bureaus: Implication for bureaucratic slack. Journal of Productivity Analysis 7:29–39
Dalen DM, Gomez-Lobo A (1997) Estimating cost functions in regulated industries under asymmetric information. European Economic Review 31:935–942
Dalen DM, Gomez-Lobo A (2001) Yardstick on the Road: Regulatory Contracts and Cost Efficiency in the Norwegian Bus Industry. Working Paper, Norwegian School of Management
Dalgaard P (2002) Introductiory Statistics with R. Springer
Davison A, Hinkley D (1997) Bootstrap Methods and Their Application. Cambridge
Debreu G (1951) The coefficient of resource utilization. Econometrica 19(3):273–292
Demsetz H (1968) Why regulate utilities? Journal of Law and Economics 11
Deprins D, Simar L, Tulkens H (1984) Measuring labor efficiency in post offices. Tech. rep., In M. Marchand, P. Pestieau, and H. Tulkens, (eds.) "The Performance of Public Enterprises: Concepts and Measurements", North Holland, pp243-267
Dorfman R, Samuelson P, Solow R (1958) Linear Programming and Economic Analysis. New York
Efron B, Tibshirani RJ (1993) An introduction to the bootstrap. Chapman & Hall
Eldén L, Witttmeyer-Koch L, Nielsen HB (2004) Introduction to numerical computation – analysis and MATLAB illustrations. Studentlitteratur, Lund
Färe R, Grosskopf S (2000) Network dea. Socio-Economic Planning Sciences pp 35–49
Färe R, Lovell C (1978) Measuring the technical efficiency of production. Journal of Economic Theory 19:150–162
Färe R, Primont D (1995) Multi-Output Production and Duality: Theory and Applications. Kluwer Academic Publishers, Boston

Färe R, Grosskopf S, Lovell CAK (1985) The Measurement of Efficiency of Production. Kluwer Nijhoff Publishing
Färe R, Grosskopf S, Lovell CAK, Yaisawatng S (1993) Derivation of shadow prices for undesirable outputs: A distance function approach. Review of Economics and Statistics 75:374–380
Färe R, Grosskopf S, Lindgren B, Ross P (1994) Productivity development in swedish hopsitals: A malmquist output index approach. In: Data Envelopment Analysis: theory, methodology, and application, Kluwer Academic Publishers, chap 13, pp 253–272
Farrell MJ (1957) The Measurement of Productive Efficiency. Journal of the Royal Statistical Society 120:253–281
Fethi M, Jackson PM, Weyman-Jones TG (2001) European airlines: a stochastic dea study of efficiency with market liberalisation. Tech. rep., University of Leicester Efficiency and Productivity Research Unit
Førsund F, Hjalmarsson L (1979) Generalized farrell measures of efficiency: An application to milk processing in swedish dairy plants. Economic Journal 89:294 – 315
Førsund F, Kittelsen S (1998) Productivity development of norwegian electricity distribution utilities. Resource and Energy Economics 20:207–224
Fox KJ (1999) Efficiency at different levels of aggregation: public vs. private sector firms. Economics Letters 65
Frandsen PE, Jonasson K, Nielsen HB, Tingleff O (2004) Unconstrained optimization. Informatics and Mathematical Modelling, Technical University of Denmark, DTU, URL http://www.imm.dtu.dk/courses/02611/uncon.pdf
Gale D (1960) The Theory of Linear Economic Models. McGraw-Hill: New York
Gill PE, Murray W, Wright MH (1981) Practical Optimization. Academic Press: London
Government TF (2007) Verordnung zum erlass und zur änderung von rechtsvorschriften auf dem gebiet der energieregulierung. Germany Teil I Nr. 55, Bundesgesetzblatt
Greene W (2008) Econometric Analysis, sixth edn. Pearson Prentice Hall
Greene WH (1990) A Gamma-Distributed Stochastic Frontier Model. Journal of Econometrics 46:141–164
Hadley G (1962) Linear Programming. Reading, Massachusetts: Addison Wesley
Halme M, Joro T, Korhonen P, Salo S, Wallenius J (1999) A value efficiency approach to incorporating preference information in data envelopment analysis. Management Science 45:103–115
Hillier FS, Lieberman GJ (2010) Introduction to Operations Research, ninth edn. McGraw-Hill
Hoff A (2007) Second stage dea: Comparison of approaches for modelling the dea score. European Journal of Operational Research 181:425–435
Jondrow J, Lovell CK, Materov IS, Schmidt P (1982) On the estimation of technical inefficiency in the stochastic frontier production function model. Journal of Econometrics 19:233–238
Joro T, Korhonen P, Wallenius J (1998) Structural comparison of data envelopment analysis and multiple objective linear programming. Management Science 44:962–970
Kalai E, Smorodinsky M (1975) Other solutions to nash's bargaining problem. Econometrica 43(3):513–519
Koopmans T (1951) Activity Analysis of Production and Allocation. Wiley, New York
Kumbhakar SC, Lovel CAK (2000) Stochastic Frontier Analysis. Cambridge University Press
Kuosmanen T (2001) Dea with efficiency classification preserving conditional convexity. European Journal of Operational Research 132:83–99
Kuosmanen T (2003) Duality theory of non-convex technologies. Journal of Productivity Analysis 20
Laffont JJ, Tirole J (1993) A Theory of Incentives in Procurement and Regulation. MIT Press
Land KC, Lovel CAK, Thore S (1993) Chance-constrained data envelopment analysis. Managerial and Decision Economics 14:541–554
Langset T (2009) Rundskriv eø 4/2009 om beregning av inntektsrammer og kostnadsnorm for 2010. (In Norwegian) NVE 2009 04925-4, The Norwegian Water Resources and Energy Directorate (NVE)
Lazear E, Rosen S (1981) Rank-order tournaments as optimum labor contracts. Journal of Political Economy 89:841–864

Lehmann EL (1983) Theory of Point Estimation. Wiley: New York
Lehmann EL (1986) Testing Statistical Hypotheses, 2nd edn. Wiley: New York
Lewin A, Morey RC (1981) Measuring the relative efficiency and output potential of public sector organizations: An application of data envelopment analysis. Journal of Policy Analysis and Information Systems 5:267–285
Littlechild S (1983) Regulation of british telecommunications' profitability: report to the secretary of state. Tech. rep., Department of Industry, London
Lovell CAK (1993) Production Frontiers and Productive Efficiency. In: Fried H, Lovell CAK, Schmidt S (eds) The Measurement of Productive Efficiency: Techniques and Applications, Oxford University Press, New York
Luenberger D (1992) Benefit functions and duality. Journal of Mathematical Economics 21:461–481
Luenberger DG (1984) Linear and Nonlinear Programming, 2nd edn. Addison-Wesley: Reading, Massachusetts
Maddala GS (1983) Limited–dependent and qualitative variables in econometrics. Cambridge University Press
Madsen K, Nielsen HB, Søndergaard J (2002) Robust subroutines for non-linear optimization. Tech. rep., Informatics and Mathematical Modelling, Technical University of Denmark, DTU, Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, URL http://www.imm.dtu.dk/~km/{F-_}package/robust.pdf
Malmquist S (1953) Index numbers and indifference curves. Trabajos de Estatistica 4:209–242
McDonald J (2009) Using least squares and tobit in second stage dea efficiency analyses. European Journal of Operational Research 197:792–798
Murrell P (2006) R Graphics. Chapman & Hall
Nalebuff BJ, Stiglitz JE (1983) Prizes and incentives: Towards a general theory of compensation and competition. Bell Journal of Economics 14:21–43
Olesen O, Petersen NC (1995) Chance constrained efficiency evaluation. Management Science 41(3):442–457
Olesen O, Petersen NC (2002) The use of data envelopment analysis with probabilistic assurance regions for measuring hospital efficiency. Journal of Productivity Analysis 17:83–109
Petersen N (1990) Data envelopment analysis on a relaxed set of assumptions. Management Science 36(3):305–314
Post GT (2001) Estimating non-convex production sets using transconcave dea. European Journal of Operational Research 131:132–142
Rao CR (1973) Linear Statistical Inference and Its Applications, 2nd edn. Wiley: New York
Rasmussen S (2010) Production Economics. The Basic Theory of Production Optimization. Springer
Resende M (2001) Relative efficiency measurement and prospects for yardstick competition in brazilian electricity distribution. Energy Policy In Press
Richmond J (1974) Estimating the efficiency of production. International Economic Review 15(515–521)
Ritter C, Simar L (1997) Pitfalls of normal-gamma stochastic frontier models. Journal of Productivity Analysis 8(2):167–182
Russell RR (1985) Measures of technical efficiency. Journal of Economic Theory 35:109–126
Russell RR (1987) On the axiomatic approach to the measurement of economic efficiency. In: Eichhorn W (ed) Measurement in Economics: Theory and Applications of Economic Indices, Physica-Verlag, Heidelberg
Russell RR (1990) Continuity of measures of technical efficiency. Journal of Economic Theory 51:255–267
Seiford LM (1994) A dea bibliography (1978–1992). In: Charnes A, Cooper W, Lewin A, (Eds) LS (eds) ata envelopment analysis: theory, methodology, and application, Kluwer Academic Publishers, pp 437–469

Shephard RW (1953) Cost and Production Functions. Princeton University Press: Princeton, New Jersey, reprint of the First Edition in 1981 by Springer Verlag, Berlin, in *Lecture Notes in Economics and Mathematical Systems*, volume 194

Shephard RW (1970) Theory of Cost and Production Functions. Princeton University Press: Princeton, New Jersey

Sheriff G (2001) Using data envelopment analysis to design contracts under asymmetric information. Tech. rep., University of Maryland

Shleifer A (1985) A theory of yardstick competition. Rand Journal of Economics 16:319–327

Silverman B (1986) Density Estimation for Statistics and Data Analysis. Chapman & Hall, London

Silvey SD (1970) Statistical Inference. Chapmann and Hall: London, reprinted with corrections 1975

Simar L, Wilson P (2000) A general methodology for bootstrapping in non-parametric frontier models. Journal of Applied Statistics 27(6):779–802

Simar L, Wilson P (2007) Estimation and inference in two-stage, semi-parametric models of production processes. Journal of Econometrics 136:31–64

Simar L, Wilson PW (1998) Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. Management Science 44(1):49–61

Smith P (1976) On the statistical estimation of parametric frontier production functions. Review of Economics and Statistics 58:238–239

Spector P (2008) Data manipulation with R. UseR!, Springer

Tavaras G (2002) A bibliography of data envelopment analysis (1978-2001). Tech. rep., Rutgers Centre of Operations Research

Thanassoulis E (2000) DEA and its use in the regulation of water companies. European Journal of Operational Research 127:1–13

Thanassoulis E, Portela M, Allen R (2004) Handbook on Data Envelopment Analysis, Kluwer Academic Publishers, chap Ch 4 Incorporating Value Judgements in DEA, pp 99–138

Tirole J (1988) The Theory of Industrial Organization. MIT Press

Tobin J (1958) Estimation of relationships for limited dependent variables. Econometrica 26(1):24–36

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

Varian HR (1992) Microeconomic Analysis, 3rd edn. New York: Norton

Venables W, Ripley B (1999) Modern Applied Statistics with S-Plus, 3rd edn. Springer

Wilson PW (1993) Detecting outliers in deterministic nonparametric frontier models with multiple outputs. Journal of Business & Economics Statistics 11(3):319–323

Wilson PW (2008) FEAR 1.0: A software package for frontier efficiency analysis with R. Socio-Economic Planning Sciences 42:247–254

Wunsch P (1995) Peer comparison and regulation: An application to urban mass transit firms in europe. PhD thesis, Department of Economics, UniversitÈ Catholique de Louvain, 182pp

Zieschang K (1984) An extended farrell technical efficiency measure. Journal of Economic Theory 33:387–396

Zuur AF, Ieno EN, Meesters E (2009) A Beginner's Guide to R (Use R). UseR!, Springer

Index