Data Envelopment Analysis

Theory and Techniques for Economics and Operations Research

SUBHASH C. RAY
University of Connecticut

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

First published 2004

Printed in the United States of America

Typefaces Times New Roman PS 10.5/13 pt. and Lucida Typewriter
System LaTeX 2ε [TB]

A catalog record for this book is available from the British Library.

Library of Congress Cataloging-in-Publication Data

Ray, Subhash C.

Data envelopment analysis : theory and techniques for economics
and operations research / Subhash C. Ray.

p. cm.

Includes bibliographical references and index.

ISBN 0-521-80256-3

1. Data envelopment analysis. 2. Production (Economic theory). I. Title.

HA31.38.R39 2004
338.4′5′01 – dc22 2003061673

Contents

1

Introduction and Overview

1.1 Data Envelopment Analysis and Economics

Data Envelopment Analysis (DEA) is a nonparametric method of measuring the efficiency of a decision-making unit (DMU) such as a firm or a public-sector agency, first introduced into the Operations Research (OR) literature by Charnes, Cooper, and Rhodes (CCR) (*European Journal of Operational Research* [EJOR], 1978). The original CCR model was applicable only to technologies characterized by constant returns to scale globally. In what turned out to be a major breakthrough, Banker, Charnes, and Cooper (BCC) (*Management Science*, 1984) extended the CCR model to accommodate technologies that exhibit variable returns to scale. In subsequent years, methodological contributions from a large number of researchers accumulated into a significant volume of literature around the CCR–BCC models, and the generic approach of DEA emerged as a valid alternative to regression analysis for efficiency measurement. The rapid pace of dissemination of DEA as an acceptable method of efficiency analysis can be inferred from the fact that Seiford (1994) in his DEA bibliography lists no fewer than 472 published articles and accepted Ph.D. dissertations even as early as 1992. In a more recent bibliography, Tavares (2002) includes 3,183 items from 2,152 different authors. Indeed, at the present moment, an Internet search for DEA produces no fewer than 12,700 entries! Parallel development of computer software for solving the DEA linear programming (LP) problems made it considerably easier to use DEA in practical applications. Apart from the LP procedures within general-purpose packages like SAS and SHAZAM, specialized packages like Integrated Data Envelopment System (IDEAS) and Data Envelopment Analysis Program (DEAP) eliminate the need to solve one LP at a time for each set of DMUs being evaluated. As a result, applying DEA to measure efficiency using a large data set has become quite routine. Unlike in Management Science where DEA became virtually an instant success, in economics, however, its welcome has been far less enthusiastic. There are three principal reasons for skepticism about DEA on the part of economists.

First, DEA is a nonparametric method; no production, cost, or profit function is estimated from the data. This precludes evaluating marginal products, partial elasticities, marginal costs, or elasticities of substitution from a fitted model. As a result, one cannot derive the usual conclusions about the technology, which are possible from a parametric functional form.

Second, DEA employs LP instead of the familiar least squares regression analysis. Whereas a basic course in econometrics centered around the classical linear model is an essential ingredient of virtually every graduate program in economics, familiarity with LP can by no means be taken for granted. In textbook economics, constraints in standard optimization problems are typically assumed to be binding and Lagrange multipliers are almost always positive. An average economist feels uncomfortable with shadow prices that become zero at the slightest perturbation of the parameters.

Finally, and most important of all, being nonstatistical in nature, the LP solution of a DEA problem produces no standard errors and leaves no room for hypothesis testing. In DEA, any deviation from the frontier is treated as inefficiency and there is no provision for random shocks. By contrast, the far more popular stochastic frontier model explicitly allows the frontier to move up or down because of random shocks. Additionally, a parametric frontier yields elasticities and other measures about the technology useful for marginal analysis.

Of the three, the first two concerns can be easily addressed. Despite its relatively recent appearance in the OR literature, the intellectual roots of DEA in economics can be traced all the way back to the early 1950s. In the aftermath of World War II, LP came to be recognized as a powerful tool for economic analysis. The papers in the Cowles Commission monograph, *Activity Analysis of Production and Resource Allocation*, edited by Koopmans (1951), recognized the commonality between existence of nonnegative prices and quantities in a Walras–Cassel economy and the mathematical programming problem of optimizing an objective function subject to a set of linear inequality constraints. Koopmans (1951) defined a point in the commodity space as efficient whenever an increase in the net output of one good can be achieved only at the cost of a decrease in the net output of another good. In view of its obvious similarity with the condition for Pareto optimality, this definition is known as the Pareto–Koopmans condition of technical efficiency. In the same year, Debreu

(1951) defined the "coefficient of resource utilization" as a measure of technical efficiency for the economy as a whole, and any deviation of this measure from unity was interpreted as a deadweight loss suffered by the society due to inefficient utilization of resources.

Farrell (1957) made a path-breaking contribution by constructing a LP model using actual input–output data of a sample of firms, the solution of which yields a numerical measure of the technical efficiency of an individual firm in the sample. In fact, Farrell's technical efficiency is the same as the distance function proposed earlier by Shephard (1953). Apart from providing a measure of technical efficiency, Farrell also identified allocative efficiency as another component of overall economic efficiency.

Linear Programming and Economic Analysis by Dorfman, Samuelson, and Solow (DOSSO) (1958) brought together the three branches of linear economic analysis – game theory, input–output analysis, and LP – under a single roof. At this point, LP came to be accepted as a computational method for measuring efficiency in different kinds of economic decision-making problems.

Farrell recognized that a function fitted by the ordinary least squares regression could not serve as a production frontier because, by construction, observed points would lie on both sides of the fitted function. He addressed this problem by taking a nonparametric approach and approximated the underlying production possibility set by the convex hull of a cone containing the observed input–output bundles. Farrell's approach was further refined by a group of agricultural economists at the University of California, Berkeley (see the papers by Boles, Bressler, Brown, Seitz, and Sitorus in a symposium volume of the Western Farm Economic Association published in 1967). In fact, a paper by Seitz subsequently appeared in *Journal of Political Economy,* one of the most prestigious and mainstream journals in economics.

Aigner and Chu (1968) retained a parametric specification of a production frontier but constrained the observed data points to lie below the function. They proposed using mathematical programming (either linear or quadratic) to fit the specified function as close to the data as possible. In a subsequent extension of this approach, Timmer (1971) allowed a small number of the observed data points to lie above the frontier in an attempt to accommodate chance variation in the data.

In a parallel strand in the literature, Afriat (1972) and Hanoch and Rothschild (1972) proposed a variety of tests of consistency of the observed data with technical and economic efficiency. One could, for example, ask whether a sample of observed input–output quantities was technically efficient.

Similarly, when input price data were also available, one could ask whether the observed firms were choosing input bundles that minimized cost. One would, of course, need to specify the technology to answer these questions. Further, the answer would depend on what form of the production technology was specified. What Afriat and Hanoch and Rothschild investigated was whether there was *any* production technology satisfying a minimum number of regularity conditions like (weak) monotonicity and convexity with reference to which the observed data could be regarded as efficient. Like Farrell (1957), they also took a nonparametric approach and used LP to perform the various tests. Although these regularity tests were designed for screening individual data points prior to fitting a production, cost, or profit function econometrically, the degree of violation of the underlying regularity conditions at an individual data point often yields a measure of efficiency of the relevant firm. Diewert and Parkan (1983) further extended the literature on nonparametric tests of regularity conditions using LP. Varian (1984) offered a battery of nonparametric tests of various properties of the technology ranging from constant returns to scale to subadditivity. Moreover, he formalized the nonparametric tests of optimizing behavior as *Weak Axiom of Cost Minimization* (WACM) and *Weak Axiom of Profit Maximization* (WAPM). More recently, Banker and Maindiratta (1988) followed up on Varian to decompose profit efficiency into a technical and an allocative component and defined upper and lower bounds on each component.

It is clear that DEA fits easily into a long tradition of nonparametric analysis of efficiency using LP in economics. In fact, in the very same year when the CCR paper appeared in *EJOR*, Färe and Lovell (1978) published a paper in *Journal of Economic Theory* in which a LP model is specified for measurement of nonradial Pareto–Koopmans efficiency.

The problem with the nonstatistical nature of DEA is much more fundamental. In fact, the lack of sampling properties of the technical efficiency of a firm obtained by solving a mathematical programming problem was recognized as a limitation of this procedure virtually right from the start. Winsten (1957), in his discussion of Farrell's paper, speculated that the frontier relationship between inputs and output would be parallel to but above the average relationship. This evidently anticipated the so-called corrected ordinary least squares (COLS) procedure that adjusts the intercept for estimating a deterministic production frontier (see Richmond [1974]; Greene [1980]) by two decades. Similarly, the production frontier was conceptualized as stochastic by Sturrock (1957), another discussant of Farrell's paper, who pointed out that the output producible from an input bundle would be subject to chance variations beyond the control of the firm and argued against using "freakishly good" results to define 100 percent efficiency.

Lack of standard errors of the DEA efficiency measures stems from the fact that the stochastic properties of inequality-constrained estimators are not well established in the econometric literature. Even in a simple two-variable linear regression with a nonnegativity constraint on the slope coefficient, the sampling distribution of the constrained estimator is a discrete–continuous type and the estimator is biased (see Theil [1971], pp. 353–4). Naturally, for a DEA model with multiple inequality constraints, the problem is far more complex and a simple solution is unlikely in the near future. At this point in time, however, there are several different lines of research underway to address this problem.

First, Banker (1993) conceptualized a convex and monotonic nonparametric frontier with a one-sided disturbance term and showed that the DEA estimator converges in distribution to the maximum likelihood estimators. He also specified F tests for hypothesis testing. Subsequently, Banker and Maindiratta (1992) introduced an additional two-sided component in the composite error term and proposed an estimation procedure of the nonparametric frontier by DEA.

Second, several researchers (e.g., Land, Lovell, and Thore [1993]) have applied chance-constrained programming allowing the inequality constraints to be violated only with a prespecified low probability.

Third, a line of research initiated by Simar (1992) and Simar and Wilson (1998, 2000) combines bootstrapping with DEA to generate empirical distributions of the efficiency measures of individual firms. This has generated a lot of interest in the profession and one may expect the standard DEA software to incorporate the bootstrapping option in the near future.

Finally, in a related but somewhat different approach, Park and Simar (1994) and Kniep and Simar (1996) have employed semiparametric and nonparametric estimation techniques to derive the statistical distribution of the efficiency estimates.

1.2 Motivation for This Book

At present, an overwhelming majority of practitioners remain content with merely feeding the data into the specialized DEA packages without much thought about whether the LP model solved is really appropriate for the problem under investigation. The more enterprising and committed researcher has to struggle through the difficult articles (many of which appeared in OR journals) in order to understand the theoretical underpinnings of the various types of LP models that one has to solve for measuring efficiency. The principal objective of this book is to deal comprehensively with DEA for efficiency measurement in an expository fashion for economists. At the same time, it seeks to provide the economic theory behind the various DEA models for the benefit of an OR/management science (MS) analyst unfamiliar with neoclassical production theory. The book by Färe, Grosskopf, and Lovell (FGL) (1994) does provide a rigorous and systematic discussion of efficiency measurement using nonparametric LP-based methods. But their persistent use of set theoretic analysis intimidates the average reader. On the other hand, the more recent book by Coelli, Rao, and Battese (1998) is, as the authors acknowledge, designed to provide a lower level bridge to the more advanced books on performance measurement.

By far the most significant book on DEA in the MS/OR strand of the literature is the recent publication by Cooper, Seiford, and Tone (2000). The authors carefully develop the different DEA models and cover in meticulous detail various mathematical corollaries that follow from the important theorems. As such, it is essential reading for one who wants to pursue the technical aspects of DEA. Designed primarily for the OR analyst, however, it understandably lacks the production economic insights behind the various models.

The present volume is designed to fill a gap in the literature by systematically relating various kinds of DEA models to specific concepts and issues relating to productivity and efficiency in economics. It may be viewed as a somewhat "higher level" bridge to the more advanced material and is meant primarily for readers who want to learn about the economic theoretical foundations of DEA at an intuitive level without sacrificing rigor entirely. This background should enable them to set up their own DEA LP models that best capture the essence of the context under which efficiency is being measured.

The chapters include numerous examples using real-life data from various empirical applications. In most cases, a typical SAS program and the output from the program are included for the benefit of the reader.

1.3 An Overview

The following is a brief outline of the broad topics and themes around which the different chapters have been developed.

Measurement of Productivity and Technical Efficiency without Price Data

Productivity and technical efficiency are two closely related but different measures of performance of a firm. They are equivalent only when the technology exhibits constant returns to scale (CRS). Chapter 2 develops the basic DEA model formulated by CCR for measurement of technical efficiency of individual firms under CRS using observed input–output quantity data. A simple transformation of the variables reduces the CCR ratio model involving a linear fractional functional programming into an equivalent LP problem. An appendix to this chapter includes a discussion of the Shephard distance function and its various properties for the interested reader. The CRS assumption is relaxed in Chapter 3, in which the BCC model applicable to technologies with variable returns to scale is presented. The maximum average productivity attained at the most productive scale size (MPSS) is compared with the average productivity at the actual scale of production to measure scale efficiency. The chapter also presents several alternative ways to determine the nature of returns to scale at an observed point. These two chapters are by far the most important in the entire volume, and a thorough grasp of the material contained in them is essential for a complete understanding of the rest of the chapters.

Chapter 4 presents various extensions to the basic DEA models considered in the earlier chapters. These include (1) the use of the graph hyperbolic distance function and the directional distance function for efficiency measurement, (2) rank ordering firms, all of which are evaluated at 100% efficiency based on DEA models, (3) identifying influential observations in DEA, and (4) a discussion of invariance properties of various DEA models to data transformation. In many situations, there are factors influencing the technical efficiency of a firm that are beyond the control of the producer. These are treated as nondiscretionary variables. One may include these variables within the constraints but not in the objective function of the DEA model. Alternatively, in a two-step procedure, they may be excluded from the DEA in the first stage but specified as independent variables in a second-stage regression model explaining the efficiency scores obtained in the first stage. Chapter 4 also considers the conceptual link between the DEA scores and the subsequent regression model in such a two-step procedure. The reader may skip this chapter at first reading and may choose to return to it at a later stage.

Pareto–Koopmans Technical Efficiency

Pareto–Koopmans technical efficiency is incompatible with unrealized output potential and/or avoidable input waste. Of course, when all outputs and inputs have strictly positive market prices, cost minimization automatically results in a Pareto–Koopmans efficient input bundle and profit maximization results in a similarly efficient input–output bundle. In the absence of market prices, however, one seeks the maximum equiproportionate increase in all outputs or equiproportionate decrease in all inputs. This is known as radial efficiency measurement. Both the CCR and BCC models fall into this category. But such an efficient radial projection of an observed input–output bundle onto the frontier does not necessarily exhaust the potential for expansion in all outputs or potential reduction in all inputs. The projected point may be on a vertical or horizontal segment of an isoquant, where the marginal rate of substitution between inputs equals zero. A different and nonradial model for efficiency measurement was first proposed by Färe and Lovell (1978) and is similar to the invariant additive DEA model.

Chapter 5 considers nonradial projections of observed input–output bundles onto the efficient segment of the frontier where marginal rates of substitution (or transformation) are strictly positive. In such models, outputs and inputs are allowed to change disproportionately.

Efficiency Measurement without Convexity

In DEA, convexity of the production possibility set is a maintained hypothesis. Convexity ensures that when two or more input–output combinations are known to be feasible, any weighted average of the input bundles can produce a similarly weighted average of the corresponding output bundles. In Free Disposal Hull (FDH) analysis, one dispenses with the convexity requirement and retains only the assumption of free disposability of inputs and outputs. FDH analysis relies on dominance relations between observed input–output bundles to measure efficiency. Chapter 6 deals with FDH analysis as an alternative to DEA and shows how FDH results in a more restricted version of the mathematical programming problem in DEA. Although not essential for an overall understanding of DEA, the material presented in this chapter helps the reader to fully appreciate the important role of the convexity assumption.

Slacks, Multiplier Bounds, and Congestion

Presence of input and/or output slacks at the optimal solution of a radial DEA model is an endemic problem. An alternative to the nonradial models considered in Chapter 5 is to ensure *a priori* that no such slacks remain at an optimal solution. The methods of Assurance Region (AR) and Cone Ratio (CR) analysis, described in Chapter 7, focus on the dual of the CCR or BCC model but put bounds on the dual variables. This ensures that the corresponding restriction in the primal problem will hold as equality. As a result, all potential for output gain and input saving is fully realized and Pareto–Koopmans technical efficiency is attained.

Underlying the horizontal or vertical segment of an isoquant or a product transformation curve is the assumption of free or strong disposability of inputs or outputs. Free disposability of inputs, for example, implies that increase in the quantity of any input without any reduction in any other input will not cause a reduction in output. One could simply leave the additional quantity of the particular input idle. In some cases, however, input disposal is costly. In agricultural production, for example, water for irrigation is an input with positive marginal productivity. If, however, excessive rain causes flooding, one needs to use capital and labor for drainage. At this stage, marginal productivity of water has become negative and the isoquant is not horizontal but upward sloping because additional quantities of other inputs are required to neutralize the detrimental effects of excessive irrigation. Along the upward rising segment of the isoquant, in the two-input case, it is possible to increase both inputs (but not only one) without reducing output. This is known as weak disposability of inputs and results in what is described as input congestion. The problem of congestion is also considered in Chapter 7.

Breakup and Merger of Firms

The production technology is super-additive if the output bundles produced individually by two firms can be produced more efficiently together by a single firm. There is an efficiency argument in favor of merger of these two firms. Similarly, in some cases, breaking up an existing firm into a number of smaller firms would improve efficiency. In economics, the question of sub-/super-additivity of the cost function and its implication for the optimal structure of an industry was investigated in detail by Baumol, Panzar, and Willig (1982). Maindiratta's (1990) definition of "size efficiency" applies the same concept in the context of DEA. Chapter 8 deals with the efficiency implications of merger and breakup of firms.

Measurement of Economic Efficiency Using Market Prices

Attaining technical efficiency ensures that a firm produces the maximum output possible from a given input bundle or uses a minimal input quantity to produce a specified output level. But no account is taken of the substitution possibilities between inputs or transformation possibilities between outputs. Full economic efficiency lies in selecting the cost-minimizing input bundle when the output is exogenously determined (e.g., the number of patients treated in a hospital) and in selecting the profit-maximizing input and output bundles when both are choice variables, as in the case of a business firm. Chapter 9 considers first the cost-minimization problem and then the profit-maximization problem in DEA. Following Farrell, the cost efficiency is decomposed into technical and allocative efficiency factors. Similarly, lost profit due to inefficiency is traced to technical and allocative inefficiency components. Chapter 9 provides the crucial link between DEA and standard neoclassical theory of a competitive firm and plays a key role in the overall development of the volume.

Nonparametric Tests of Optimizing Behavior

Chapter 10 presents some of the major tests for optimizing behavior in producer theory existing in the literature. This chapter considers Varian's Weak Axiom of Cost Minimization and its relation to a number of related procedures. Diewert and Parkan (1983) and Varian (1984) define an outer and an inner approximation to the production possibility set based on the quantity and price information about inputs and outputs of firms in a sample. These yield the lower and upper bounds of various efficiency measures. The material presented here is primarily of a methodological interest and may be skipped by a more empirically motivated reader.

Productivity Change over Time: Malmquist and Fisher Indexes

Caves, Christensen, and Diewert (CCD) (1982) introduced the Malmquist productivity index to measure productivity differences over time. Färe, Grosskopf, Lindgren, and Roos (FGLR) (1992) developed DEA models that measure the Malmquist index. There is a growing literature on decomposition of the Malmquist index into separate factors representing technical change, technical efficiency change, and scale efficiency change. Apart from the Malmquist index, Chapter 11 also shows the measurement and decomposition of the Fisher index using DEA. In light of the increasing popularity of this topic, this chapter is highly recommended even to the average reader.

Stochastic Data Envelopment Analysis

By far the most serious impediment to a wider acceptance of DEA as a valid analytical method in economics is that it is seen as nonstatistical, not distinguishing inefficiency from random shocks. Although a satisfactory resolution of the problem is not at hand, efforts to add a stochastic dimension to DEA have been made along several different lines. Chapter 12 presents Banker's F tests, Chance-Constrained Programming, Varian's statistical test of cost minimization, and bootstrapping for DEA as various major directions of research in this area. Of these, bootstrapping appears to be most promising and is becoming increasingly popular. Chapter 12 is essential reading for every serious reader.

Beyond the standard CCR and BCC DEA models, the choice of topics that are to be included in a standard reference textbook is largely a matter of preference of the author. In the present case, topics that are more directly related to neoclassical production economics have been included. Others, like multi-criterion decision making (MCDM) and goal programming – although by no means less important in the context of DEA – have been excluded. Readers interested in these and other primarily OR/MS aspects of DEA should consult Cooper, Seiford, and Tone (2000) for guidance.