International Series in
Operations Research & Management Science

Joe Zhu *Editor*

Data Envelopment Analysis

A Handbook of Empirical Studies and Applications

Operations Research Management Science

Springer

International Series in Operations Research & Management Science

Volume 238

More information about this series at http://www.springer.com/series/6161

Joe Zhu
Editor

Data Envelopment Analysis

A Handbook of Empirical Studies and Applications

Springer

Editor
Joe Zhu
International Center for Auditing and Evaluation
Nanjing Audit University
Nanjing, P.R., China

School of Business
Worcester Polytechnic Institute
Worcester, MA, USA

ISSN 0884-8289 ISSN 2214-7934 (electronic)
International Series in Operations Research & Management Science
ISBN 978-1-4899-7682-6 ISBN 978-1-4899-7684-0 (eBook)
DOI 10.1007/978-1-4899-7684-0

Library of Congress Control Number: 2015959708

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

Preface

This handbook complements *Handbook on Data Envelopment Analysis* (eds, W.W. Cooper, L.M. Seiford, and J. Zhu, 2011, Springer), *Data Envelopment Analysis*: *A Handbook of Modeling Internal Structures and Networks* (eds, W.D. Cook and J. Zhu, 2014, Springer), and *Data Envelopment Analysis*: *A Handbook of Models and Methods* (ed. J. Zhu, 2015, Springer). Data envelopment analysis (DEA) is a "data-oriented" approach for evaluating the performance of a set of entities called decision-making units (DMUs) whose performance is categorized by multiple metrics. These performance metrics are classified or termed as inputs and outputs under DEA. Although DEA has a strong link to production theory in economics, the tool is also used for benchmarking in operations management, where a set of measures is selected to benchmark the performance of manufacturing and service operations. In the circumstance of benchmarking, the efficient DMUs, as defined by DEA, may not necessarily form a "production frontier," but rather lead to a "best-practice frontier" (Cook, Tone, and Zhu, 2014).

Over the years, we have seen a variety of DEA empirical applications. This handbook aims to compile state-of-the-art empirical studies and applications using DEA. It includes a collection of 18 chapters written by DEA experts.

Chapter 1, by Chen, Gregoriou, and Rouah, examines the performance of chief executive officers (CEOs) of US banks and thrifts. The authors find evidence that best-practice CEOs who have a DEA efficiency score of one are rewarded with higher compensation compared to underperforming CEOs who have a DEA efficiency score greater than one. They also find DEA efficiency score to be a highly significant predictor of CEO compensation.

Chapter 2, by Yu and Chen, is dedicated to describe the network operational structure of transportation organizations and the relative network data envelopment analysis model. Route-based performance evaluation, environmental factors, undesirable outputs, and multi-activity framework are incorporated into their application.

Chapter 3, by Hu and Chang, demonstrates how to use different types of DEA models to compute the total-factor energy efficiency scores with an application to energy efficiency.

Chapter 4, by Growitsch, Jamasb, Müller, and Wissner, explores the impact of incorporating customers' willingness to pay for service quality in benchmarking models on cost efficiency of distribution networks.

Chapter 5, by Volz, provides a brief review of previous applications of DEA to the professional baseball industry followed by two detailed applications to Major League Baseball.

Chapter 6, by Cummins and Xie, examines efficiency and productivity of US property-liability (P-L) insurers using DEA. The authors estimate pure technical, scale, cost, revenue, and profit efficiency over the period 1993–2011. Insurers' adjacent year total-factor productivity changes, and their contributing factors are also investigated.

Chapter 7, by Premachandra, Zhu, Watson, and Galagedera, presents a two-stage network DEA model that decomposes the overall efficiency of a decision-making unit into two components and demonstrates its applicability by assessing the relative performance of 66 large mutual fund families in the USA over the period 1993–2008.

Chapter 8, by Basso and Funari, presents a comprehensive review of the literature of DEA models for the performance assessment of mutual funds along with an empirical application on real market data, considering different risk measures. The authors consider different holding periods, which include both a period of financial crisis and one of financial recovery.

Chapter 9, by Hwang and Chang, discusses the management strategies formulation of the international tourist hotel industry in Taiwan based on the efficiency evaluation. The result of this chapter provides useful information for future business management needs of managers and can be served as valuable reference to the relevant authority of tourism.

Chapter 10, by Chen, Zhu, Yu, and Noori, presents a novel use of the two-stage network DEA to evaluate the sustainable product design performances. A two-stage network DEA model is developed for sustainable design performance evaluation with an "industrial design module" and a "bio design module." Test results show that sustainable design does not need to mean compromise between traditional and environmental attributes.

Chapter 11, by Chen and Ang, highlights limitations of some DEA environmental efficiency models, including directional distance function and radial efficiency models, under weak disposability assumption and various return-to-scale technologies. The empirical results show that the directional distance function and radial efficiency models may generate spurious efficiency estimates, and thus it must be with caution when they are used for decision support.

Chapter 12, by Thanassoulis, De Witte, Johnes, Johnes, Karagiannis, and Portela, reviews applications of DEA in secondary and tertiary education, focusing on the opportunities that this offers for benchmarking at institutional level.

Chapter 13, by Sexton, Comunale, Higuera, and Stickle, measures the relative performance of New York State school districts in the 2011–2012 academic year and provided detailed alternative improvement pathways for each district.

Chapter 14, by Iyer and Grewal, provides an introductory chapter as prelude to Chap. 15, by Grewal, Levy, Mehrotra, and Sharma, and Chap. 16, by Grewal, Iyer, Kamakura, Mehrotra, and Sharma. Both Chaps. 15 and 16 provide detailed applications of DEA in marketing.

Chapter 17, by O'Donnell, shows how to decompose a new total-factor productivity index that satisfies all economically relevant axioms from index theory with an application to US agriculture.

This handbook concludes with Chap. 18, by Liu, Lu, and Lu. This unique study conducts a DEA research front analysis. The large amount of DEA literature makes it difficult to use any traditional qualitative methodology to sort out the matter. Thus, this study applies a network clustering method to group the literature through a citation network established from the DEA literature over the period 2000–2014. The findings are helpful in many ways, including identifying coherent topics or issues addressed by a group of research articles in recent years.

I hope that this handbook, along with other aforementioned DEA handbooks, can serve as a reference for researcher and practitioners using DEA and as a guide for further development of DEA. I am indebted to the many DEA researchers worldwide for their continued effort in pushing the DEA research frontier. Without their work, many of the DEA models, approaches, and applications would not exist. I would like to thank the support from the Priority Academic Program Development of Jiangsu Higher Education Institutions in China.

Worcester, MA, USA Joe Zhu
October 2015

References

Cook WD, Tone K, Zhu J (2014) Data envelopment analysis: prior to choosing a model. Omega 44:1–4

Cook WD, Zhu J (2014) Data envelopment analysis: a handbook of modeling internal structures and networks, Springer

Cooper WW, Seiford LM, Zhu J (2011) Handbook on data envelopment analysis, 2nd edn. Springer, New York

Zhu J (2015) Data envelopment analysis: a handbook of models and methods. Springer

Contents

Contributors

Sheng Ang School of Management, University of Science and Technology of China, Hefei, People's Republic of China

Antonella Basso Department of Economics, Ca' Foscari University of Venice, Venice, Italy

Tzu-Pu Chang Department of Finance, National Yunlin University of Science and Technology, Yunlin, Taiwan

Te-Yi Chang Graduate Institute of Travel and Tourism Management, National Kaohsiung University of Hospitality and Tourism, Kaohsiung, Taiwan

Li-Hsueh Chen Department of Transportation Science, National Taiwan Ocean University, Keelung, Taiwan

Chialin Chen Department of Business Administration, National Taiwan University, Taipei, Taiwan, R.O.C.

Chien-Ming Chen Nanyang Business School, Nanyang Technological University, Singapore, Singapore

Yao Chen International Center for Auditing and Evaluation, Nanjing Audit University, Nanjing, People's Republic of China

Manning School of Business, University of Massachusetts at Lowell, Lowell, MA, USA

Christie Comunale, Ph.D. College of Business, Stony Brook University, Stony Brook, NY, USA

J. David Cummins Temple University, Philadelphia, PA, USA

Stefania Funari Department of Management, Ca' Foscari University of Venice, Venice, Italy

Don U. A. Galagedera Department of Econometrics and Business Statistics, Monash Business School, Monash University, Melbourne, VIC, Australia

Greg N. Gregoriou School of Business and Economics, State University of New York (Plattsburgh), Plattsburgh, NY, USA

Dhruv Grewal Babson College, Babson Park, MA, USA

Christian Growitsch University of Hamburg, Hamburg, Germany

Michael Shane Higuera, Ed.D. College of Business, Stony Brook University, Stony Brook, NY, USA

Jin-Li Hu Institute of Business and Management, National Chiao Tung University, Hsinchu, Taiwan

Shiuh-Nan Hwang Department of Business Administration, Ming Chuan University, Taipei, Taiwan

Gopalkrishnan R. Iyer Center for Services Marketing & Management, Florida Atlantic University, Boca Raton, FL, USA

Tooraj Jamasb Durham University Business School, Durham, UK

Jill Johnes Huddersfield University Business School, University of Huddersfield, Huddersfield, UK

Geraint Johnes Lancaster University Management School, Lancaster University, Lancaster, UK

Wagner A. Kamakura Rice University, Houston, TX, USA

Giannis Karagiannis Department of Economics, University of Macedonia, Thessaloniki, Greece

Michael Levy Babson College, Babson Park, MA, USA

John S. Liu Graduate Institute of Technology Management, National Taiwan University of Science and Technology, Taipei, Taiwan

Louis Y. Y. Lu College of Management, Yuan Ze University, Taoyuan, Taiwan

Wen-Min Lu Department of Financial Management, National Defense University, Taipei, Taiwan

Anuj Mehrotra University of Miami, Miami, FL, USA

Christine Müller Department Energy Markets and Energy Regulation, WIK, Bad Honnef, Germany

Hamid Noori Laurier School of Business, Wilfrid Laurier University, Waterloo, ON, Canada

Conceição S. Portela Católica Porto Business School, Porto, Portugal

I. M. Premachandra Department of Accountancy and Finance, School of Business, University of Otago, Dunedin, New Zealand

Fabrice Douglas Rouah Desautels Faculty of Management, McGill University, Montreal, QC, Canada

Thomas R. Sexton, Ph.D. College of Business, Stony Brook University, Stony Brook, NY, USA

Arun Sharma University of Miami, Miami, FL, USA

Kelly Stickle College of Business, Stony Brook University, Stony Brook, NY, USA

Emmanuel Thanassoulis Aston Business School, Aston University, Birmingham, UK

Brian D. Volz Department of Economics and Global Studies, Assumption College, Worcester, MA, USA

John Watson Department of Banking and Finance, Monash Business School, Monash University, Melbourne, VIC, Australia

Matthias Wissner Department Energy Markets and Energy Regulation, WIK, Bad Honnef, Germany

Kristof De Witte Leuven Economics of Education Research, Faculty of Economics and Business, KU Leuven, Leuven, Belgium

Top Institute for Evidence Based Education Research, Maastricht University, Maastricht, The Netherlands

Xiaoying Xie California State University—Fullerton, Fullerton, CA, USA

M.-M. Yu Department of Transportation Science, National Taiwan Ocean University, Keelung, Taiwan

Jiun-Yu Yu Department of Business Administration, National Taiwan University, Taipei, Taiwan

Joe Zhu International Center for Auditing and Evaluation, Nanjing Audit University, Nanjing, P.R., China

School of Business, Worcester Polytechnic Institute, Worcester, MA, USA

Chapter 1
Efficiency Persistence of Bank and Thrift CEOs Using Data Envelopment Analysis

Yao Chen, Greg N. Gregoriou, and Fabrice Douglas Rouah

Abstract We examine the performance of chief executive officers (CEOs) of U.S. banks and thrifts. We apply Data Envelopment Analysis (DEA) to measure the performance of CEOs on a yearly basis over the 1997–2004 period, and find evidence that best-practice CEOs who have a DEA efficiency score of one are rewarded with higher compensation compared to under-performing CEOs who have a DEA efficiency score greater than one. We find DEA efficiency score to be a highly significant predictor of CEO compensation, even after adjusting for firm size. In addition, we find that DEA efficiency scores of CEOs have decreased over the observation period. We also find that best-practice CEOs tend to be persistent on a yearly basis, but we find little evidence of multi-period persistence. The results of this study can serve as a benchmark for CEOs wishing to evaluate their performance relative to their peers, and as a new measure of CEO performance.

Keywords Chief executive officers (CEOs) • Data envelopment analysis (DEA) • Performance • Compensation • Thrifts

---

This chapter is based upon Chen, Y., G. N. Gregoriou, and F. D. Rouah (2009), "Efficiency persistence of bank and thrift CEOs using data envelopment analysis", Computers and Operations Research, Vol. 36, Issue 5, 1554–1561. with permission from Elsevier. Professor Yao Chen thanks the Priority Academic Program Development of the Jianhsu Higher Education Institutions (China) for their support of this work.

Y. Chen (✉)
International Center for Auditing and Evaluation, Nanjing Audit University, Nanjing, People's Republic of China

Manning School of Business, University of Massachusetts at Lowell, Lowell, MA 01845, USA
e-mail: Yao_Chen@uml.edu

G.N. Gregoriou
School of Business and Economics, State University of New York (Plattsburgh), 101 Broad Street, Plattsburgh, NY 12901, USA
e-mail: Greg.Gregoriou@plattsburgh.edu

F.D. Rouah
Desautels Faculty of Management, McGill University, 1001 Sherbrooke Street West, Montreal, QC, Canada H3A 1G5
e-mail: Fabrice.Rouah@mail.mcgill.ca

J. Zhu (ed.), *Data Envelopment Analysis*, International Series in Operations Research & Management Science 238, DOI 10.1007/978-1-4899-7684-0_1

1.1 Introduction

This chapter examines how the compensation of CEOs of U.S. banks and thrifts depends on the performance of CEOs in allocating the resources of their firm, over the 1997–2004 period. The mandate of the CEO of a large banking institution can be very daunting, especially considering the wide range of products banks and thrifts now offer, which range from credit cards, loans, and investments in stocks, bonds and mutual funds. Compensation is higher and more varied than ever before, with more than 90 % of CEOs at large U.S. companies receiving stock options according to Watson Wyatt Worldwide (2006). Not surprisingly, salaries and bonuses have come under a great deal of scrutiny by shareholders, activists, and the media, since many firms have performed sub-optimally despite the lucrative compensation packages being awarded to CEOs and other firm executives. There is, however, some evidence that the compensation structure of bank CEOs is somewhat different than that of CEOs in other sectors, and that stock compensation in the banking industry is lower, on average, than in other sectors (Houston and James 1995).

While in theory compensation packages are intended to incite firm executives (CEOs) to act in the best interest of their shareholders, it is often argued that in practice, compensation packages are excessively high and are not related to performance. In numerous academic studies, the relationship between performance and compensation is explored only with traditional regression models. Our study goes one step further by examining the yearly performance of bank CEOs using data envelopment analysis (DEA), and whether this performance is persistent over time. The CEO performance in the current study is defined by the DEA model used and is characterized by multiple performance measures.

DEA was developed by Charnes et al. (1978) to assess the efficiency of decision making units (DMU) that have multiple inputs and outputs. DEA has advantages over traditional parametric techniques—such as regression—that are worth pointing out. Regression models assume that the relationship between executive compensation and performance is linear and subject to arbitrary fluctuations and predicted values from such a model are constructed relative to the average performance of all CEOs. In contrast, DEA examines each CEO uniquely, by generating individual performance (efficiency) scores that are relative to the entire sample under investigation. Misspecification is a recurring problem in regression analysis. However, misspecification is not a concern with DEA models, since DEA creates a best practices frontier based on peer comparisons within the sample. Furthermore, studies such as that by Gregoriou and Rouah (2003) find common factors among variables that are often linked with CEO compensation in banks, which can lead to problems of multicollinearity in regression models. DEA on the other hand can handle multiple performance measures (called inputs and outputs) in a single mathematical model without the need for the specification of tradeoffs among multiple measures related to CEO performance. DEA has been demonstrated to be a valuable instrument for performance evaluation and benchmarking (see, for example, Zhu (2014) and Cooper et al. 2004)).

Data on CEO compensation is collected yearly, so applying longitudinal analysis is difficult, especially when few data points are available. Many studies of CEO compensation are therefore cross-sectional. In this study, we examine the relationship between compensation and performance, but also investigate whether best-practice CEOs who have a DEA score of one are rewarded with higher compensation, and attempt to identify longitudinal persistence in performance as measured in the DEA efficiency scores. We propose that DEA efficiency can serve as an additional metric for measuring CEO performance. For some CEOs, best performance may occur only sporadically, while others may be DEA efficient year after year. Establishing persistence among CEOs can help identify star executives, those that understand how their banking institution works and the measures needed to provide a greater revenue stream and a higher net income.

The chapter is organized as follows. The next section provides a literature review, followed by the sections describing the data and methodology. This is followed by the empirical results and a discussion of our findings. The final section concludes.

1.2 Literature Review

CEO compensation has been the subject of academic research since the early 1990s. While the results of existing studies are varied, many find a strong and significant association between CEO compensation and bank performance (Sigler and Porterfield 2001). Hall and Liebman (1998) and Lambert and Larcker (1987) find a significant relationship between CEOs and firm value, but Kerr and Bettis (1987) do not come across any such relationship. Other studies, such as those by Agarwal and Mandelker (1987), Core et al. (1999), Guay and Core (1999), Joyce (2001) and Murphy (1998), for example, uncover a weak correlation between executive compensation and firm performance. There is further evidence by Sigler and Porterfield (2001) that an increase in bank revenue leads to an increase in CEO compensation, while McIntyre and Rao (1993) discover an association between compensation and return to shareholders. In addition, Bosworth et al. (2003) observe that large banks are more efficient than small ones, a finding supported by Bliss and Rosen (2001), who further detect a link between firm size and CEO compensation. On the other hand, Lambert and Larcker (1987) and Banker and Datar (1989) conclude that there is no definite relationship between accounting measures and compensation or between market returns and compensation. Early and highly quoted papers by Murphy (1985) and Coughlan and Schmidt (1985) detect significant relationships between stock market performance and CEO compensation. These studies, however, are conducted before the deregulation of U.S. banks and thrifts by the Reagan administration in the early 1980s.

One difficulty faced by these and other studies is that CEO performance is somewhat difficult to define and measure, and that compensation packages frequently create agency[1] problems. Hence, some authors have proposed compensation plans that could help induce CEOs to act in the best interest of their firm and alleviate these problems. Gibbons and Murphy (1990) suggest that CEO compensation be based on a stock market index that whereby the stock option's exercise price is linked to the performance of the index. A landmark study by Jensen and Murphy (1990) identify that changes in incentive contracts and accounting profits align the interest of CEOs with those of the shareholders. Furthermore, Hall and Murphy (2000) find a strong correlation between unexercised options of CEOs and bank performance relative to their peer group. Although, Hall (2001) further finds that stock options are not an efficient tool for motivating CEOs to enhance firm performance, Ofek and Yermack (2000) demonstrate that CEOs unload their shares of their compensation package immediately after receiving the restricted stock options.[2] CEOs may find that it is optimal to cash options after the time restriction, but in so doing may signal to shareholders that their performance is mediocre and could therefore be replaced. This is consistent with the finding of Hall (2001) that many stock options are not very effective in motivating CEOs to improve firm performance. As suggested by Jesuthasan et al. (2000) firms could alleviate agency problems by linking stock option prices to realized future growth in the industry, which would encourage CEO performance. Chen et al. (2009) suggest that efficiency could serve as an additional measure of CEO performance, and that compensation could be linked to how efficiently the CEO is able to use the resources of the firm to generate revenue and maximize profit.

1.3 Data and Methodology

To examine the performance of CEOs of U.S. banks and thrifts, we use the SNL Executive Compensation database, which contains yearly data collected from nearly 3000 private and public banking institutions. We use data covering the period January 1997 to December 2004. SNL obtains secondary data from industry sources such as Business Week or Forbes, and Securities and Exchange Commission (SEC) documentation available on company websites. Our dataset initially consists of 3213 bank and thrift CEOs, but we only examine CEOs that are at their present position at December 2004, thus reducing the number of CEOs to 283. A handful of CEOs were dropped each year because of insufficient data. Table 1.1 presents yearly descriptive statistics on the CEOs in our sample, including option-adjusted compensation, and the revenue, expenses, and income of their banking

[1] Agency problems are present when information is asymmetric, i.e., information known only by insiders of the firm.

[2] For example, a stock option granted to a CEO having an exercise price of $5 is considered worthless when the stock trades at $2, but a restricted stock option with the same exercise price and trades at $2 then it has only lost 40 % of its value.

Table 1.1 Descriptive statistics of CEOs and firms, by year

	Number of observation	CEO compensation ($000)		Firm revenue ($M)		Firm expense ($M)		Firm income ($M)	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
1997	220	507	818	89	359	51	209	22	89
1998	233	588	927	102	395	57	225	24	88
1999	249	685	1018	122	414	70	232	30	108
2000	260	797	1826	165	560	93	316	41	144
2001	274	973	2,21	310	1545	173	925	66	272
2002	271	1002	1786	262	975	136	482	68	254
2003	270	1068	2049	280	1067	148	535	75	300
2004	268	1207	2342	322	1150	171	585	93	358
All	2045	869	1764	212	931	116	509	54	231

Number of observations in the sample, and mean and standard deviation (S.D.) of CEO compensation (in thousands), firm revenue (in millions), firm expenses (in millions), and firm income (in millions), from 1997 to 2004 yearly, and for all years combined (last row)

institutions. It shows both CEO compensation and firm income to have doubled over the observation period, while firm expense has not risen nearly as dramatically. At first glance, it appears that the increase in CEO compensation is justified, in light of the large rise in the income of their firms and the ability of the CEOs to keep expenses under control.

Data envelopment analysis is a relatively new technique that has been applied to measure efficiency in various areas of research, including hedge funds (Gregoriou and Zhu 2005) and banks (Seiford and Zhu 1999; Paradi et al. 2004). We apply DEA to relate a set of banking inputs, which are the resources available to the CEO, to a set of outputs, which are the performance measures of the CEO's firm. DEA uses these inputs and outputs to calculate an efficiency score for each CEO. The most efficient CEOs are those that use the least amount of input to produce the greatest amount of output. CEOs achieving an efficiency score of 1.0 are deemed efficient and are located on the best-practice frontier.

The best way to evaluate the long-run performance of CEOs is to mix CEO-specific variables with firm variables measured over the long term. Traditionally, CEOs of major public corporations have a long-term approach to profit maximization and have a moral responsibility to maximize shareholder wealth through good judgment and proper strategic vision, which can help avoid the deterioration of corporate profits. Since the majority of investors are long-term investors, they are likely to be concerned about the performance of CEOs over long time periods. By evaluating the persistency of CEO performance over time, we address these concerns and recommend that efficiency scores can serve as a metric for CEO compensation. Many studies have used assets, deposits, revenue, net income and number of employees as factors related to bank performance efficiency (see for example, Seiford and Zhu 1999). The mix of input variables used in this study demonstrates that option-adjusted compensation of CEOs and firm variables are able to identify how those CEOs that can generate profits for their firms.

We reproduce the definitions of inputs and outputs from the www.snl.com database. The first input we use is option adjusted compensation which includes base salary, annual bonus, options, and other compensation, as recorded by the Securities and Exchange Commission in the bank's filings. Total option adjusted compensation includes the estimated value of options granted to the CEO during the period. The second input is non-interest expense, which represents the total non-interest expense excluding nonrecurring items and minority interest expense. Noninterest expenses largely consist of data processing costs, occupancy charges, and personnel making these areas important for CEOs to use cost cutting measures.

The third input is the sum of all assets (total assets) which includes short-term assets such as cash, investments, inventory, and receivables and long-term assets, for example, property and equipment owned by the bank at period-end identify how aggressive the CEO's expansion strategy in this area. The fourth input is the deposits of the bank which consist of the total interest and non-interest-bearing deposits at period end, including passbook, checking, NOW, time and any other deposit in a federally insured bank or thrift. This input gages if the CEO is successful in his strategy to attract new clients to increase growth of the bank's dollar assets.

The fifth input is the number of total full service locations at period end and identifies if the CEO has undertaken a project to continue and open-up full service branches year after. Finally, the sixth and final input is the number of employees on a full-time equivalent basis at period end which should detect if there are more full time employees at branches due to increased business.

We select two outputs firm revenue and net income as the two outputs since we take into account the direct relationship between input variables and output variables. Outputs are the result of processing inputs, and measure how efficiently a CEO has attained his or her goals. Revenue includes nonrecurring revenue and is net of interest expense for banks, thrifts, lenders, federal home loan banking system, investment companies, asset managers and brokers/dealers.

The second output is net income, which includes total revenue, net of total expense, income taxes, minority interest, extraordinary items and other after-tax-adjustments. These input and output variables will help us to determine which CEOs are good at minimizing expenses or maximizing the revenue and net income of their firms—those that are skillful cost-cutters and concerned with the bottom line.

We select these inputs and outputs because CEOs typically have a tendency to cost cut, increase revenue, and increase profits. Moreover, management stability should be reflected in their DEA efficiency scores, so that best-practice (or DEA efficient) CEOs should stay with their firms longer than under-performing CEOs.

In this study, the CEOs are the DMUs. Suppose the jth CEO ($j = 1, \ldots, n$) uses i inputs ($i = 1, \ldots, m$) to produce r outputs ($r = 1, \ldots, s$). We define x_{ij} to be the quantity of input i that CEO j uses to produce the quantity y_{rj} of output r. Each CEO uses m different inputs to generate s different outputs. We presume that $x_{ij} \geq 0$,

$y_{rj} \geq 0$, and that each CEO has at least one positive input value and one positive output value. In this study, $n = 277$, $m = 3$, and $s = 2$.

DEA optimization uses the values x_{ij} and y_{rj} to select values of input and output weights for a CEO. The efficiency score θ^* of a CEO is the solution to the following problem

$$\begin{aligned} &\max\theta \\ &\text{subject to} \\ &\sum_{j=1}^{n}\lambda_j x_{ij} \leq x_{i0} \qquad i = 1,2,\ldots,m; \\ &\sum_{j=1}^{n}\lambda_j y_{rj} \geq \theta y_{r0} \qquad r = 1,2,\ldots,s; \\ &\lambda_j \geq 0 \qquad\qquad\quad j = 1,2,\ldots,n. \end{aligned}$$

In these constraints x_{i0} is the amount of input i used by CEO_0, y_{r0} is the amount of output r produced by CEO_0, and CEO_0 is the CEO under evaluation in the optimization. This model used in this chapter is the output-oriented constant returns to scale (CRS) model developed by Charnes et al. (1978). We use an output oriented model since we are assessing CEO performance. Using the CRS model controls for size and we compare each CEO with all other CEOs in the sample. The output-oriented CRS model is equivalent to the input-oriented CRS model. Therefore, the use of the above model tries to determine which CEOs are good at minimizing expenses or maximizing the revenue and net income of their firms.

In the above model, the optimal value θ^* is the (DEA) efficiency for a CEO under evaluation. If $\theta^* = 1$, the CEO is deemed efficient or a best-practice, but if $\theta > 1$, the CEO is inefficient or under-performing.

1.3.1 CEO Compensation, Efficiency, and Persistence

We propose that efficient CEOs are rewarded with higher compensation. To test this assertion, we run a regression model of compensation on DEA score, but controlling for firm size, since the CEOs of larger firms also earn larger salaries. Firm expenses, income, and revenue can all be used to proxy firm size. Because of the high degree of correlation between these measures,[3] however, we use only firm expense. Expense is chosen because one of the primary concerns of managers is to reduce the expenses of their firm. If efficiency plays a role in explaining CEO

[3] The correlation between expenses, revenue, and income is at least 0.90 each year.

compensation, then the DEA score should come out significant in regressions, even after controlling for firm expenses. Hence we run the model

$$y_{it} = \alpha_t + \beta_{1t}\,DEA_{it} + \beta_{2t}\,Exp_{it} + \beta_{3t}\,Exp^2{}_{it} + \varepsilon_{it}$$

where

y_{it} = compensation earned by CEO i during year t,
DEA_{it} = DEA score of CEO i during year t,
Exp_{it} = Expenses of firm i during year t,

and where, for each year t, α_t, β_{1t}, β_{2t} and β_{3t} are regression coefficients and ε_{it} is an error term. We except $\beta_{1t} < 0$ since our hypothesis is that efficiency and compensation are positively related. We also expect $\beta_{2t} > 0$ since we expect compensation to be positively related to the size of the firm. We also include the square of expenses because we expect decreasing returns to scale for compensation with firm size. Hence we expect $\beta_{3t} < 0$.

To test whether efficiency is persistent from 1 year to the next, we define five states of DEA scores as follows: State 1: $\theta^* = 1$; State 2: $1 < \theta^* \leq 1.3$; State 3: $1.3 < \theta^* \leq 1.5$; State 4: $1.5 < \theta^* \leq 1.7$; and State 5: $\theta^* > 1.7$. Note that State 1 corresponds to DEA efficiency. We then define the transition matrix P of dimension 5×5, with elements p_{ij} denoting the conditional probability of being in state j in 1 year, given that the CEO was in state i the previous year

$$p_{ij} = \Pr\left[\text{CEO in state } j \text{ in year } t+1 \;\middle|\text{CEO in state } i \text{ in year } t\right].$$

The model we are proposing is a Markov chain with a finite number of states. Hence, in building this transition matrix, we make the assumption that yearly DEA scores are Markovian, so that the probability of the CEO being in state j during year $t+1$ depends only on what state the CEO was in year t, and not in the state during years prior to year t. We also assume that DEA scores evolve yearly and remain constant in any given year.

The transition probabilities are estimated by grouping transitions from all CEOs and during all years together, and calculating the proportion of observed transitions from state i to state j. If CEOs are persistent from 1 year to the next, this would be reflected in a large value of p_{55}. We also obtain the steady state probabilities of the matrix, which helps to assess the long-run probability of CEOs being efficient. These probabilities $p = (p_1, p_2, p_3, p_4, p_5)$ are the row elements of the matrix obtained by multiplying P with itself infinitely many times. The vector p can be obtained as the solution of the linear system $pP = p$.

As an alternate way to evaluate persistence, we regress current yearly DEA scores on past yearly scores. Hence assume that current yearly DEA scores are driven by the autoregressive relationship, as specified by the following AR(1) model

$$DEA_{it} = \alpha + \beta\,DEA_{i,t-1} + v_{it}$$

where DEA_{it} = DEA score for CEO i in year t,

$DEA_{i,t-1}$ = DEA score for CEO i in year $t-1$,
$v_{it} = \varepsilon_{it} - \varphi v_{i,t-1}$,
$\varepsilon_{it} \sim N(0, \sigma)$,

Efficiency persistence is indicated by a positive regression coefficient, namely $\beta > 0$.

Finally, to evaluate how well the DEA scores compares against a benchmark for evaluating compensation we calculate the correlation between DEA scores and (*i*) observed compensation, (*ii*) compensation predicted by our regression model, and (*iii*) firm income as a proportion of firm revenue. If DEA is a good predictor of compensation, these correlations should all be large and significant.

1.4 Empirical Results

Table 1.2 presents the frequency distribution of the DEA score, θ*, by year and over the entire observation period. The mean score was 1.49 over the period, with low variability. We find that the proportion of efficient CEOs to decrease slightly. In 1997, 10 % (22/220) were efficient, but by 2004 this had dropped to 7.5 % (20/268). We also find a slight increase in the yearly median efficiency scores, which suggests that banking CEOs became less efficient over the observation period.

Table 1.3 presents the yearly simple DEA efficiency scores θ* of each CEO in our sample. For each CEO, we calculate the number of times the CEO reaches

Table 1.2 Frequency distributions of simple DEA scores, by year

Efficiency score	1997	1998	1999	2000	2001	2002	2003	2004	All
1	22	23	24	20	20	20	21	20	170
1–1.3	45	55	56	44	47	27	36	30	340
1.3–1.4	28	36	39	27	47	11	37	31	256
1.4–1.5	36	40	33	31	44	20	37	34	275
1.5–1.6	36	28	42	31	41	26	49	52	305
1.6–1.7	24	23	20	33	28	26	35	32	221
1.7–1.8	12	14	13	24	23	42	23	26	177
>1.8	17	14	22	50	24	99	32	43	301
Number of scores	220	233	249	260	274	271	270	268	2045
Mean	1.42	1.41	1.42	1.52	1.45	1.67	1.49	1.54	1.49
Median	1.44	1.40	1.41	1.53	1.44	1.71	1.51	1.53	1.49
Standard Dev.	0.26	0.26	0.27	0.30	0.26	0.35	0.27	0.30	0.30
Min	1	1	1	1	1	1	1	1	1
Max	2.18	2.24	2.22	2.36	2.13	2.57	2.40	2.59	2.60

Simple DEA scores for CEOs, from 1997 to 2004 yearly, and for all years combined. The top part presents the frequency distribution of DEA scores, and the second part presents the number of scores, the mean score, median, standard deviation, minimum, and maximum

Table 1.3 Simple DEA efficiency scores of CEOs, by year

	DEA score								Proportion
CEO name	1997	1998	1999	2000	2001	2002	2003	2004	Efficient
Conrad Hanson	1	1	1	1	1	1	1	1	1
Nicholas W. Lazares	1	1	1	1	1	1	1	1	1
Salomon Levis	1	1	1	1	1	1	1	1	1
Richard N. Wayne	1	1	1	1	1	1	1	1	1
Daniel H. Cuddy	1	1	1	1	1	1	1	–	1
Ernest S. Rady	1	–	1	1	1	1	1	1	1
Jerry A. Grundhofer	–	–	–	–	1	1	1	1	1
John A. Nash	–	–	–	1	–	1	–	–	1
Stephen Lange Ranzini	–	–	–	–	–	1	1	1	1
William I. Miller	1	1	1	–	–	–	1	1	1
Lawrence A. Collett	1	1	1.07205	1	1	1	1	1	0.875
Marion O. Sandler	1.00002	1	1	1	1	1	1	1	0.875
Michael R. Melvin	1	1	1	1	1	1	1	1.02714	0.875
James H. Blanchard	1.03314	.	1	1	1	1	1	1	0.85714
George W. Haligowski	1.02595	1	1	1	1	1	1	1.01347	0.75
Herbert M. Sandler	1	1	1.00003	1.00003	1	1	1	1	0.75
Robert J. Glickman	1	1	1	1	1	1.06834	1.06606	1	0.75
Victor J. Galan	1	1	1	1.12547	1	1.09579	1	1	0.75
William A. Osborn	1	1	1	1	1	1	1.1922	1.2532	0.75
Joseph R. Ficalora	1.0855	1.0119	1	1	1.09216	1	1	1	0.625

Yearly simple DEA score for the top 20 CEOs in our sample, ranked according to the proportion of the years under observation that the CEO achieved an efficient score ($\theta^* = 1$)

Table 1.4 Transition matrix of simple DEA scores

	Score in year $t+1$						
		1	(1, 1.3]	(1.3, 1.5]	(1.5, 1.7]	>1.7	Total
Score in year t	1	0.77 (111)	0.21 (30)	0.01 (1)	0.0 (0)	0.01 (1)	1.0 (143)
	(1, 1.3]	0.08 (25)	0.61 (182)	0.22 (66)	0.06 (19)	0.03 (8)	1.0 (300)
	(1.3, 1.5]	0.00 (1)	0.10 (46)	0.51 (232)	0.26 (119)	0.12 (53)	1.0 (451)
	(1.5, 1.7]	0.00 (1)	0.02 (8)	0.23 (101)	0.45 (194)	0.29 (127)	1.0 (431)
	>1.7	0.00 (0)	0.01 (3)	0.08 (32)	0.29 (116)	0.62 (250)	1.0 401
	Total	138	269	432	448	439	1726

Transition matrix of DEA scores from year t, to year $t+1$, for all CEOs combined. Entries are transition probabilities, with the number of transitions in parentheses

efficiency, and divide by the number of years the CEO appears in our sample, which produces the proportion of years under observation for which the CEO is efficient. Hence, C. Hanson, N.W. Lazares, S. Levis, and R.N. Wayne each appear every year in the sample, and are a best-practice each year, so their proportions are each 100 %. These are the CEOs that show the best possible efficiency persistence in the sample. E.S. Rady is deemed as best-practice each year under observation, so his proportion is also 100 %, but he is missing during 1998 in our sample. The next three CEOs are efficient roughly 88 % of the time, and the proportion decreases rapidly thereafter. The other CEOs, however, are on the best-practice frontier less than 50 % of the years under observation. Our sample consists of 283 CEOs, but there is missing data in some of the years. The number of CEOs with non-missing data range from 220 in 1997 to 274 in 2001. Over the entire sample of 283 CEOs, 241 (85 %) were never able to achieve best-practice in any of the years. The rest (42) achieved a score of $\theta^* = 1$ at least once during the eight years.

Given these results, it is useful to investigate whether the presence of efficient scores originate from the same CEOs continuing to be efficient from 1 year to the next, indicating the ability of CEOs to persist in their best-practice performance. In Table 1.4 we present a transition matrix of scores from 1 year to the next, for all years and all CEOs in our sample, and using the same classifications for θ^* that appear in Table 1.2. The matrix shows a clear tendency for DEA scores to persist from one period to the next, as indicated by the large proportions of entries along the diagonal. For example, a CEO with a score of $\theta^* = 1$ in 1 year has a 77 % chance (111 transitions out of a possible 143) of being best-practice in the following year, and only a 1 % chance of having a score $\theta^* > 1.7$. Unfortunately, persistence is also evident among the CEOs with low scores, but is not as strong. A CEO with a score of $\theta^* > 1.7$ in 1 year has a 62 % chance of having the same low score in the next year. The steady state probabilities of this transition matrix are $p_1 = 0.0451$, $p_2 = 0.1067$, $p_3 = 0.2419$, $p_4 = 0.2931$, and $p_5 = 0.3133$. Hence, the long-run

Table 1.5 Auto regression of current DEA scores on past scores

Variable	Estimate	Standard error	t-statistic	p-value
Intercept (α)	0.286	0.023	12.5	<0.0001
Past DEA score (β)	0.821	0.015	55.3	<0.0001
AR(1) parameter (ϕ)	−0.347	0.023	−15.4	<0.0001

Yearly regression of current yearly DEA score on past yearly score. Entries are regression coefficients, standard errors, t-statistics, and p-values. The R-squared is 0.64 and the Durbin-Watson test statistic is 2.17

prospect for CEO efficiency is poor. Indeed, the long-run chance of a CEO being efficient ($\theta^* = 1$) is only 4.5 %, while long-run probability that a CEO is in the lowest state ($\theta^* > 1.7$) is 31.3 %. Clearly, the steady state probabilities increase with worsening DEA score. This means that in the long run, CEOs have a much greater change of being highly inefficient that efficient.

As an alternative method to examine yearly persistence, we regress current yearly DEA scores on past yearly scores. The results are presented in Table 1.5. They suggest persistence in yearly scores, as indicated by positive and significant β coefficients for past scores. This implies that good scores tend to be followed by good scores, but that bad scores tend to be followed by bad scores. For example, a CEO with a score of $\theta^* = 1$ during 1 year can be expected to have a score of roughly $\theta^* = 1.13$ in the second year ($0.347 + 0.780 \times 1$). Similarly, a poor score of $\theta^* = 2$ can be expected to be followed by a score of only $\theta^* = 1.9$ the following year. The positive relationship between current and past yearly scores is illustrated in Fig. 1.1. The figures shows a positive relationship between current and past scores, which is consistent with the argument of score persistence and illustrates that low scores in 1 year tend to be followed by low scores the next year, and similarly for high scores.

Table 1.6 presents the results of the regression model of compensation on DEA score and firm expense. As expected, $\beta_{2t} > 0$ and $\beta_{3t} < 0$ for all years, implying a concave relation between compensation and firm size. The coefficient for DEA score (β_{1t}) is negative for all years, but fails to achieve significance at the 10 % level or better in 1997 and 2000. This implies that during most years, a low DEA score was associated with a large compensation, irrespective of the size of the firm. The relationship between DEA score and compensation is especially strong in 2001 and 2002. After 2002, however, the relationship is weaker. From 2002 onwards, it appears that efficiency did not play as important a role in determining CEO compensation as it did in previous years. The adjusted R^2 from the regressions are all reasonably high.

Finally, in Table 1.7 we present the yearly correlations of the DEA score with CEO compensation, with compensation predicted from the regression model, and with the proportion of income to revenue. Table 1.7 indicates that all of the correlations are negative, and all are highly significant. Hence, DEA and compensation are negatively related. This implies that high compensation is awarded to efficient CEOs. Income/Revenue and DEA are also negatively related, which

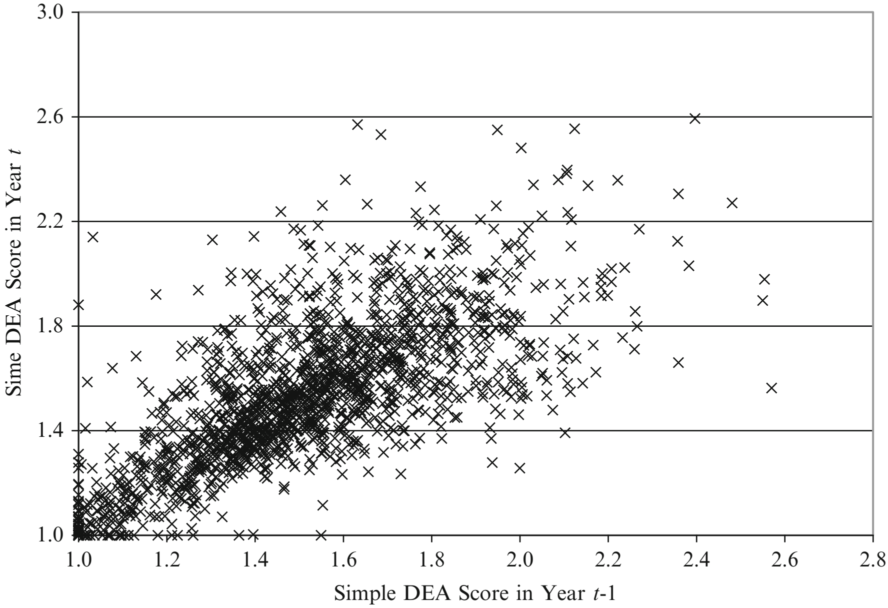


Fig. 1.1 Scatter plot of current simple DEA scores on past scores, 1997 to 2004. Scatter plot of yearly simple DEA score in year t (y-axis) and in year $t-1$ (x-axis)

Table 1.6 Regression of compensation, by year

	Intercept	DEA score × 100	Expense ($M)	Expense^2 ($M^2)	
	α_t	β_{1t}	β_{2t}	β_{3t}	Adjusted R^2
1997	6872**	−26.4	54.3***	−0.0176***	0.37
1998	9336***	−39.6**	43.7***	−0.0068***	0.52
1999	12,348***	−57.4***	44.3***	−0.0075***	0.56
2000	4459	−11.9	70.6***	−0.0114***	0.65
2001	16,800***	−83.2***	36.2***	−0.0015***	0.82
2002	24,643***	−110.9***	33.0***	−0.0025***	0.57
2003	15,027***	−57.9*	30.7***	−0.0008	0.56
2004	16,462***	−61.9*	29.9***	0.0009	0.62
All	13,577***	−57.1***	34.8***	−0.0014***	0.62

Yearly regression of CEO compensation (in hundreds of thousands) on DEA score (multiplied by 100), expense of the firm (in millions), and expenses squared (in millions squared). Entries are regression coefficients. ***, **, and * refer to coefficients significant at the 1, 5, and 10 %, respectively. The last row is the regression for the entire 1997 to 2004 period

implies that DEA can serve as a alternate benchmark of CEO compensation. Finally, predicted compensation and DEA are negatively related, which implies that our model does a good job at capturing the relationship between compensation and DEA score. This suggests that DEA scores perform well when compared to a benchmark for CEO compensation.

Table 1.7 Correlation analysis, by year

	Correlation of DEA Score with		
Year	Compensation	Predicted compensation	Income/Revenue
1997	−0.33***	−0.53***	−0.44***
1998	−0.34***	−0.46***	−0.54***
1999	−0.43***	−0.57***	−0.54***
2000	−0.37***	−0.45***	−0.47***
2001	−0.38***	−0.42***	−0.52***
2002	−0.48***	−0.63***	−0.43***
2003	−0.32***	−0.42***	−0.42***
2004	−0.34***	−0.43***	−0.54***
All	−0.32***	−0.41***	−0.44***

Yearly correlation of DEA score with compensation, with compensation predicted from the regression model of Table 1.6, and with the proportion of firm income to firm revenue. Entries are Pearson correlations for each year 1997–2004, and the last row are correlations for all years combined. ***, **, and * refer to correlations significant at the 1, 5, and 10 %, respectively

1.5 Discussion

CEOs with efficiency scores equal to unity ($\theta^* = 1$) lie on the best-practice frontier, and no other CEOs are able to generate better output level given the same number of inputs, or to generate the same output level using less amount of inputs. CEOs with the highest scores are assumed to possess the greatest amount of inefficiency. The magnitude of their scores can help identify how much effort a CEO would need to achieve best-practice. For example, a score of $\theta^* = 1.5$ implies that the CEO is 67 % efficient at using inputs to produce outputs. The CEO would need to diminish inputs by one third to be considered best-practice. Hence, many CEOs that attain an efficiency score near unity would likely need to make only minor corrections to their inputs to be considered efficient. But CEOs with scores well above $\theta^* = 1.5$ are notably far from the best-practice frontier These CEOs would need to put considerable effort into their input modifications to attain efficiency. But even CEOs with high scores may be able to attain best-practice performance if they can reduce their inputs while increasing their outputs. In summary, DEA provides a realistic representation of each CEOs degree of underperforming performance, and can provide a valuable gauge for CEOs hoping to attain best performance.

Our AR(1) regression model of current DEA scores on past DEA scores does not restrict predicted DEA scores to be one or greater. Hence, in theory, it is possible to obtain nonsensical scores less than one. None of our predicted scores fell below one, however.

In our investigation of yearly performance persistence, we find a probability of $p_{11} = 0.77$ that a CEO represents best-practice in 1 year can achieve best-practice in the next year. The long term probability of being best-practice, however, is much lower ($p_1 = 0.0451$), and multi-period persistence is much less likely. Indeed, in

Table 1.3 we find only four CEOs to have maintained their best-practice status year after year during the entire examination period. Hence, while some CEOs may be able to repeat their best performance from 1 year to the next, most are unable to maintain their best-practice status over longer time periods.

In our yearly regressions, we find evidence that best-practice performance is rewarded with higher compensation, but most for years prior to 2003. From 2003 onward, the relationship between DEA score and compensation is weaker. We find also that many CEOs are under-performing in terms of minimizing expenses and maximizing revenue and net income, which implies that they may not be justifying their large compensation packages.

1.6 Conclusion

This Chapter uses DEA to identify best-practice CEOs and to examine whether or not their compensation is warranted. We find evidence that DEA efficiency plays a role in explaining CEO compensation. We find little evidence of long-term performance persistence, however, so the large compensation packages paid out to CEOs year after year do not seem justified in most cases. The protest frequently put forward by shareholders activists, the media, and the public that CEOs often earn hefty compensation even when their firms incur large losses, seems defensible. Yet we find that a small number of CEOs do maintain their best-practice performance year after year. The efficiency score produced by DEA can serve to benchmark CEO performance, allowing CEOs to compare themselves to their peers in the banking industry, and helping shareholders and other agents decide whether the compensation packages of their CEOs are reasonable.

References

Agarwal A, Mandelker G (1987) Managerial incentives and corporate investment and financing decisions. J Finance 42(4):823–838

Banker R, Datar S (1989) Sensitivity precision and linear aggregation of signals for performance evaluation. J Account Res 27(1):21–39

Bosworth W, Mehdian SM, Vogel T (2003) Executive compensation and medium-sized bank holding companies in the United States. Am Bus Rev 21(1):91–99

Bliss R, Rosen R (2001) CEO compensation and bank mergers. J Financ Econ 61(1):107–138

Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision making units. Eur J Oper Res 2(6):429–444

Chen Y, Gregoriou GN, Rouah FD (2009) Efficiency persistence of bank and thrift CEOs using data envelopment analysis. Comput Oper Res 36(5):1554–1561

Cooper WW, Seiford LM, Zhu J (2004) Data envelopment analysis: history, models and interpretations. In: Cooper WW, Seiford LM, Zhu J (eds) Handbook on data envelopment analysis. Kluwer Academic, Boston, pp 1–39 (Chapter 1)

Core J, Holthausen RW, Larcker DF (1999) Corporate governance. Chief executive officer compensation and firm performance. J Financ Econ 51(3):371–406
Coughlan AT, Schmidt RM (1985) Executive compensation managerial turnover and firm performance: an empirical investigation. J Account Econ 5(7):43–66
Gibbons R, Murphy KJ (1990) Relative performance evaluation from chief executive officers. Ind Labor Relat Rev 43(1):30–52
Gregoriou GN, Rouah F (2003) An examination of CEO compensation of U.S. banks and thrifts using non-traditional performance measures. J Financ Serv Mark 7(3):246–257
Gregoriou GN, Zhu J (2005) Evaluating hedge fund and CTA performance: data envelopment analysis approach. Wiley, Hoboken
Guay W, Core J (1999) The use of equity grants to manage optimal equity incentive levels. J Account Econ 28(1):151–184
Hall B, Liebman JB (1998) Are CEOs really paid like bureaucrats. Quart J Econ 113(3):653–691
Hall B, Murphy K (2000) Stock options for undiversified executives. Working Paper, NBER, Cambridge
Hall B (2001) What you need to know about stock options. In: Kohn A, Zehnder E, Case J, Pfeffer J, Nicoson RD (eds) Harvard business review on compensation with Alfred Rappaport. Harvard Business School, Cambridge
Houston JF, James C (1995) CEO compensation and bank risk: is compensation in banking structured to promote risk taking? J Monet Econ 36(2):405–431
Jensen M, Murphy K (1990) CEO incentives: it's not how much you pay, but how. J Appl Corporate Finance 3(3):225–265
Jesuthasan R, Todd E, Barrett A (2000) The total performance equation. Financial Executive 16(4):41–44
Joyce WB (2001) Return and reward: bank performance and CEO compensation. Am Bus Rev 19(2):93–99
Kerr J, Bettis RA (1987) Board of directors top management compensation and shareholder returns. Acad Manag 30(4):645–664
Lambert RA, Larcker DF (1987) An analysis of the use of accounting and market measures of performance in executive compensation. J Account Res 25(S):85–125
McIntyre JE, Rao VK (1993) The effects of institutional ownership on bank CEO compensation. J Bus Econ Perspect 41(2):200–208
Murphy KJ (1985) Corporate performance and managerial remuneration: an empirical analysis. J Account Econ 7(2):11–41
Murphy KJ (1998) Executive compensation. Working Paper, USC Department of Economics, Los Angeles
Ofek E, Yermack D (2000) Taking stock equity-based compensation and the evolution of managerial ownership. J Finance 55(3):1367–1384
Paradi JC, Vela S, Yang Z (2004) Assessing bank and bank branch performance. In: Cooper WW, Seiford LM, Zhu J (eds) Handbook on data envelopment analysis. Kluwer, Boston, pp 349–400
Seiford LM, Zhu J (1999) Profitability and marketability of the top 55 U.S. commercial banks. Manag Sci 45(9):1270–1288
Sigler KJ, Porterfield R (2001) CEO compensation: its link to bank performance. Am Bus Rev 19(2):110–114
Watson Wyatt Worldwide (2006) Realties of executive compensation. Washington, DC
Zhu J (2014) Quantitative models for performance evaluation and benchmarking: DEA with spreadsheets, 3rd edn. Springer, New York

Chapter 2
Assessment of Transportation Performance: A Network Structure

Ming-Miin Yu and Li-Hsueh Chen

Abstract Performance measurement is a popular activity of organizations in the transportation sector. Various studies on the performance of transportation organizations with the utilization of data envelopment analysis models have been common. However, based on the unstorable characteristics of transportation services, conventional data envelopment analysis models are not suitable, and then network data envelopment analysis models are proposed. This chapter is dedicated to describe the network operational structure of transportation organizations and the relative network data envelopment analysis model. In order to be closer to real operational situations, four operational characteristics, which are route-based performance evaluation, environmental factors, undesirable outputs, multi-activity framework, are discussed and incorporated into the network data envelopment analysis model, respectively.

Keywords Transportation • Network DEA • Route-based performance evaluation • Environmental factors • Undesirable outputs • Multi-activity framework

---

Part of the material in this chapter is adapted from Yu, M.M. and Fan, C.K., 2006 "Measuring the cost effectiveness of multimode bus transit in the presence of accident risks", Transportation Planning and Technology, 29(5), 383–407, Yu, M.M. and Lin, E.T.J., 2008 "Efficiency and effectiveness in railway performance using a multi-activity network DEA model", Omega, 36, 1005–1017, Yu, M.M., 2008 "Measuring the efficiency and return to scale status of multimodel bus transit—evidence form Taiwan's bus system", Applied Economics Letters, 15, 647–653, Yu, M.M. and Fan, C.K., 2009 "Measuring the performance of multimode bus transit: A mixed structure network DEA model", Transportation Research Part E, 45, 501–515, Yu, M.M., 2010 "Assessment of airport performance using the SBM-NDEA model", Omega, 38, 440–452, Yu, M.M. and Chen, P.C., 2011 "Measuring air routes performance using a fractional network data envelopment analysis model", Central European Journal of Operations Research, 19, 81–98, Yu, M.M., 2012 "Performance assessment of transport services with the ERM-NDEA model: evidence from a domestic airline in Taiwan", Transportation Planning and Technology, 35 (7), 697–714, with permission from Taylor & Francis, Elsevier Science and Springer.

M.-M. Yu (✉) • L.-H. Chen
Department of Transportation Science, National Taiwan Ocean University,
2, Pei-Ning Road, Keelung 20224, Taiwan
e-mail: yumm@mail.ntou.edu.tw

J. Zhu (ed.), *Data Envelopment Analysis*, International Series in Operations Research & Management Science 238, DOI 10.1007/978-1-4899-7684-0_2

2.1 Introduction

The performance measures of the delivery of the primary services of transportation organizations have been the traditional subject of whatever performance studies were made in the past. There are many ways to scrutinize performance in the transportation sector. In early periods, the usually used measures of performance are ratio indicators, such as vehicle hours per employee, vehicle kilometers per active vehicle, passengers per revenue vehicle hour, and revenue vehicle hours per dollar operating cost (Mackie and Nash 1982; Lee 1989; Fielding 1992). Ratio analysis typically involves the use of a number of performance indicators which consider only a subset of inputs used by a decision-making unit (DMU) and sometimes only a subset of outputs. In single-input single-output contexts, a partial measure of performance is a meaningful, easy to use measure of performance.

However, this is not the case where multiple inputs and/or outputs are involved (Hensher 1992). To the extent that a DMU may increase performance with respect to one input at the expense of reducing the performance of other inputs, the difficulty stems from the fact that each partial measure of performance reflects only one input and one output level, and it is also difficult to portray the overall gains/ losses in performance (Thanassoulis et al. 1996). Furthermore, it could provide a misleading indication of overall performance when considered in isolation. In recent years, various studies on the theoretical and empirical measurement of performance in the transportation sector with the utilization of the data envelopment analysis (DEA) model have been generated by researchers. There is a large stream of literature on a single-stage DEA. In a regularly studied situation within this context, it is assumed that a transportation organization's inputs are transformed from a single operation process into their final outputs. Some of those studies focus on production efficiency (e.g., Tulkens 1993; Oben 1994; Kerstens 1996; Nolan et al. 2001; Cowie 2002; Karlaftis 2003; Graham 2008), while some are interested in the measurement of operational efficiency (e.g., Tofallis 1997; Cowie and Asenova 1999; Adler and Golany 2001; Boame 2004; Yu 2007), and others invested both in a single model (e.g., Viton 1998; McMullen and Noh 2007).

While evaluating the performance in the transportation sector, it is worth noting that, unlike the production and consumption processes of the manufacturing sector, a transportation service cannot be stored, and therefore the output consumed (the final output), such as passenger-km, may vary considerably from the output produced (the intermediate output), such as vehicle-km, in a transportation system. Specifically, the consumed services occur concurrently with the produced services. If the produced output is not consumed, it is lost (Tomazinis 1975) (e.g., if a bus runs during the period at half capacity, the bus system cannot store the other half of its inventory (Karlaftis 2004)). This perishability of the produced services and the fact that only a proportion of the produced services are actually consumed is often neglected in performance measures of transportation organizations (Borger et al. 2002). If these unique unstorable characteristics of transportation services are justified, then it is vitally important to obtain valid estimates of performance of transportation organizations that include them. Hence, an adequate performance measurement for a transportation organization should consider the network structure that services are produced and consumed concurrently, and interactions in this structure.

In addition, other operational issues, such as route-based performance evaluation, environmental factors, undesirable outputs, multi-activity framework, etc., will also impact the assessment of performance in the transportation sector. In order to construct a more reasonable performance measurement for transportation organizations, these four issues mentioned above will also be explored and incorporated into the network structure.

The remainder of this chapter is organized as follows. In the second section, we describe the transportation performance; the specifications of the network DEA model in transportation appear in the third section; in the fourth section, we explore other issues for transportation applications; the fifth section provides three examples; and concluding remarks are given in the final section.

2.2 Transportation Performance

Since transportation services cannot be stored, the output consumption may be substantially different from the output production. For instance, an airline uses aircraft, employees, and fuel to provide service products, flights, and seat-miles, which are produced and sold to passengers concurrently. Once the service products are not consumed (that is, seats are not sold), they are wasted. So service products function as intermediate inputs (the intermediate outputs in the production process) and used internally in consumption process. To accommodate unstorable characteristics, Fielding et al. (1985) introduced three performance indicators for a transit system: cost efficiency, service effectiveness, and cost effectiveness. They defined cost efficiency as the ratio of outputs to inputs, service effectiveness as the ratio of consumption to outputs, and cost effectiveness as the ratio of consumption to inputs. Hence, cost effectiveness is the integration of cost efficiency and service effectiveness measures. This transit performance concept is portrayed in Fig. 2.1.

However, the definition of "cost efficiency" used by Fielding et al. (1985) could cause some confusion, because, in the economic theory and DEA context, cost efficiency is defined by the product of technical efficiency and allocative efficiency. If the input factor prices are not available, it would be more appropriate to use the terms of production efficiency, service effectiveness and operational effectiveness instead of cost efficiency, service effectiveness and cost effectiveness, respectively.

Most studies about performance measurement used separate models to measure the interrelated processes, and evaluate sub-process efficiency independently (Chu et al. 1992; Viton 1998; Nolan et al. 2002; Lan and Lin 2003, 2005; Karlaftis 2004; Chiou and Chen 2006). They distinguished the production process from the consumption process, from which one can gain more insight into the firms' operations.

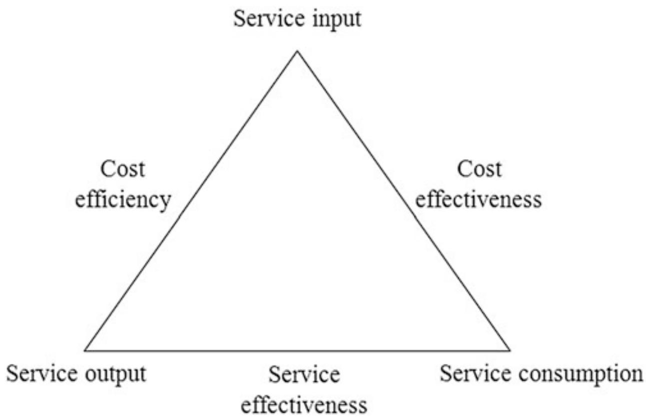


Fig. 2.1 Transit performance concept

However, since outputs are consumed concurrently with their production, measuring the performance of the transportation organizations using two models is likely to be unreasonable. In addition, these models mentioned above assume different production technologies without interacting each other and cannot deal formally with intermediate products. It ignores effects of the inter-relationship between sub-processes and then yields an incomplete version of operational performance measurement (Sheth et al. 2007). In any realistic situation, the transportation sector has a feature of unstorable series, which means that intermediate products are presented both in production and consumption processes. Usually, the feature within a transportation organization's operation should take into account all the complex and interrelated flows between these two processes. Assuming that transportation frequencies are given by a particular schedule for serving their passengers, inefficiency occurs when the actual level of input consumption, for a given level of provided capacity (e.g., frequencies and/or seat-miles), exceeds the optimal level of input requirement as specified by the production function. This observed production inefficiency, however, does not mean service ineffectiveness, since a transportation organization could search for better ways to maximize its ridership to raise its service effectiveness. In other words, service effectiveness may be seen as how a transportation organization efficiently transforms capacity provided to ridership in the consumption process. In making performance comparisons, they must take into account the multistage representation of the technology, otherwise the performance measures would reflect not merely differences in efficiency but also the relative efficiency by which individual processes and the whole operation system are operating. In addition, for a transportation organization which is obliged to provide a stable timetable in a given time period, it implies that if the predetermined timetable is violated, then the violation may result in the waste/decrease of input costs and the loss/gain of consumed outputs with respect to some referenced efficient transportation organizations since the changes in the timetable may increase or reduce cost and/or passengers may feel comfortable/uncomfortable

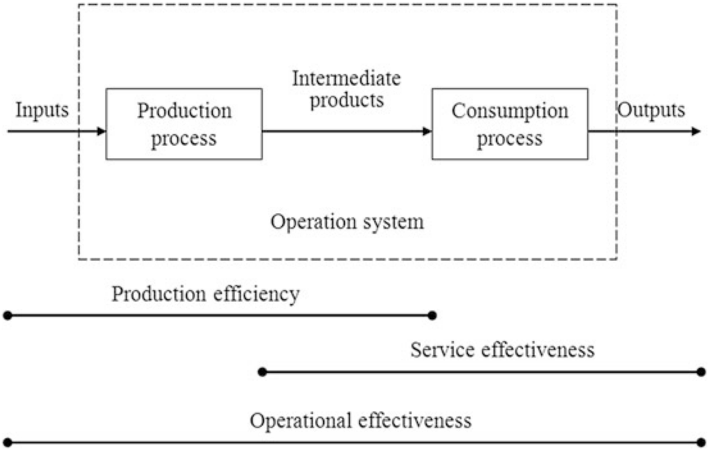


Fig. 2.2 Performance evaluation in the network structure

using it. Hence, a transportation organization possesses a network structure including a set of interdependent technologies in the whole operational process. By separating the effects of the complex and interrelated technologies, we can explore if the source of observed performance differs. Identification of such sources is essential to the implementation of operational policies and management strategies designed to improve performance. Therefore, it seems more realistic and reasonable to use a unified network model to estimate the performance of transportation organizations. This performance evaluation in network structure is shown in Fig. 2.2.

2.3 Network Data Envelopment Analysis in Transportation

Traditionally, DEA has treated each DMU as a "black box" by considering only the inputs consumed and final outputs produced by this "black box" (Färe and Grosskopf 2000). However, in most real situations, the DMUs may perform several different functions and can also be separated into different components in series. In such situations, some components play important roles in producing outputs through the use of intermediate outputs obtained from their previous components. In this case, the conventional DEA model cannot impose restrictions on the inter-relationships among intermediate products when measuring the DMU's overall performance together with that of its components. If this "black box" consists of a set of sub-units which are connected serially, then such an approach provides no insights regarding the inter-relationships among the components' inefficiencies and cannot provide specific process guidance to DMU managers to help them improve the DMU's efficiency.

In transportation organizations, the operational process in a DMU usually contains two processes in which some outputs produced in former process are used as inputs in a latter process. Färe and Grosskopf (1996, 2000) proposed a network DEA model for measuring performance of those DMUs with multiple processes. The object of this proposed method was to provide a solution to deal with a weakness, which treats the operational process as a "black box", in the conventional DEA model. In order to represent production and consumption processes in a transportation organization's operating technology, a network DEA model based on the directional distance function proposed by Luenberger (1992) is constructed as below.

We denote inputs for the production process by $x^P \in R_+^A$. Here inputs x are employed in the production process (P) to produce intermediate outputs, $m^{(P,\,C)} \in R_+^B$, where (P, C) represents the intermediate output of P flowing into the consumption process (C). Intermediate outputs from the production process act as intermediate inputs to the consumption process. The intermediate products are produced in production and consumed in consumption processes concurrently, resulting in final outputs $y^C \in R_+^D$. To formulate a network DEA model, we need to introduce intensity variables z_j^P and z_j^C, $j = 1, \ldots, J$, for production and consumption processes of each DMU j, respectively. Hence, the network DEA model has a production possibility set and a consumption possibility set, A^P, and A^C, which can be defined as follows:

$$A^P = \left\{ \left(x^P,\ m^{(P,\,C)} \right) :\ m^{(P,\,C)}\ can\ be\ produced\ from\ x^P \right\}, \tag{2.1}$$

$$A^C = \left\{ \left(m^{(P,\,C)},\ y^C \right) :\ y^C\ can\ be\ produced\ from\ m^{(P,\,C)} \right\}. \tag{2.2}$$

If A^P is the smallest set which satisfies the convexity, the constant returns to scale, free disposability, and minimum extrapolation postulates (Tsai and Mar Molinero 2002), subject to the condition that each input–output observations $\left(x^P,\ m^{(P,\,C)}\right) \in A^P$, then the input set in the production process, $P^P(m^{(P,\,C)})$, for each $m^{(P,\,C)}$ can be defined as $P^P\left(m^{(P,\,C)}\right) = \left\{x^P : \left(x^P,\ m^{(P,\,C)}\right) \in A^P\right\}$. Similarly, the output set in the consumption process, $P^C(m^{(P,\,C)})$, for each $m^{(P,\,C)}$ can be defined as $P^C\left(m^{(P,\,C)}\right) = \left\{y^C : \left(m^{(P,\,C)},\ y^C\right) \in A^C\right\}$.

An overall network operational possibility set in terms of the input and output set is defined as follows:

$$T^N\left\{\left(x^P,\ m^{(P,C)},\ y^C\right) : \sum_{j=1}^{J} z_j^P x_{aj}^P \le x_a^P,\ a = 1, \ldots, A,\right.$$

$$\sum_{j=1}^{J} z_j^P m_{bj}^{(P,C)} \ge m_b^{(P,C)},\ b = 1, \ldots, B,$$

$$\sum_{j=1}^{J} z_j^C y_{dj}^C \ge y_d^C,\ d = 1, \ldots, D, \tag{2.3}$$

$$\sum_{j=1}^{J} z_j^C m_{bj}^{(P,C)} \le m_b^{(P,C)},\ b = 1, \ldots, B,$$

$$\left. z_j^P \ge 0,\ z_j^C \ge 0,\ j = 1, \ldots, J\right\}$$

We introduce two functions: $\beta_k^P(x^P,\ m^{(P,C)})$ and $\beta_k^C(m^{(P,C)},\ y^C)$, which provide measures of how efficient a firm k is in production process and consumption process, respectively. The efficiency score of each part could be calculated as follows:

$$\begin{aligned}\vec{D}\left(x^P,\ m^{(P,C)}\right) &= \beta_k^P\left(x^P,\ m^{(P,C)}\right)\\ &= \max\left\{\beta_k^P : \left(1-\beta_k^P\right)x^P \in P^P\left(m^{(P,C)}\right),\ \beta_k^P \ge 0\right\},\end{aligned} \tag{2.4}$$

$$\begin{aligned}\vec{D}\left(m^{(P,C)},\ y^C\right) &= \beta_k^C\left(m^{(P,C)},\ y^C\right)\\ &= \max\left\{\beta_k^C : \left(1+\beta_k^C\right)y^C \in P^C\left(m^{(P,C)}\right),\ \beta_k^C \ge 0\right\}.\end{aligned} \tag{2.5}$$

For an illustration of the network performance measurement, we choose to evaluate firm k relative to the network technology (2.3) by means of a directional distance function. The objective function of the network model is taken as the form:

$$Max\ \beta^k = w_k^P \beta_k^P + w_k^C \beta_k^C, \tag{2.6}$$

where β_k^P and β_k^C are the performance scores of production and consumption processes, respectively; w_k^P and w_k^C are positive numbers which represent the relative importance of these processes respectively, and $w_k^P + w_k^C = 1$.

In the network DEA model, we can identify these two sub-technologies. Hence, (2.6) is subject to these following constraints:

The production process consists of

$$\sum_{j=1}^{J} z_j^P x_{aj}^P \le \left(1-\beta_k^P\right)x_{ak}^P,\ a = 1, \ldots, A, \tag{2.6.1}$$

$$\sum_{j=1}^{J} z_j^P m_{bj}^{(P,\,C)} \geq m_{bk}^{(P,\,C)}, \;\; b = 1, \ldots, \;\; B, \tag{2.6.2}$$

$$\beta_k^P \geq 0, z_j^P \geq 0, \;\; j = 1, \ldots, J. \tag{2.6.3}$$

The consumption process is given by,

$$\sum_{j=1}^{J} z_j^C m_{bj}^{(P,\,C)} \leq m_{bk}^{(P,\,C)}, \;\; b = 1, \ldots, \;\; B, \tag{2.6.4}$$

$$\sum_{j=1}^{J} z_j^C y_{dj}^C \geq \left(1 + \beta_k^C\right) y_{dk}^C, \;\; d = 1, \ldots, \;\; D, \tag{2.6.5}$$

$$\beta_k^C \geq 0, \;\; z_j^C \geq 0, \;\; j = 1, \ldots, J, \tag{2.6.6}$$

The network directional distance function in (2.6) is zero if and only if the transportation organization's production process is technically efficient, and its consumption process is simultaneously serviced effectively. However, its value is greater than zero if and only if the transportation organization is technically inefficient in at least one of the two processes. The network DEA model has several attractive features compared to the conventional one. In particular, it provides individual managers with specific information regarding the sources of inefficiency within their DMUs.

2.4 Other Issues for Transportation Applications

In order to resemble the real operational characteristics of transportation organizations, besides the network structure of transportation services, other operational issues must be considered. In this section, we mention four issues that transportation organizations often confront, but not all are included. These four issues are:

- Route-based performance evaluation
- Environmental factors
- Undesirable outputs
- Multi-activity framework

2.4.1 Route-Based Performance Evaluation

Most studies measure the performance of transportation organizations from a whole-company perspective. They treat individual firms as individual DMUs. However, different transportation organizations may operate different routes, such as operational routes vary among different shipping companies or airlines, even in the same country. A whole-company perspective may lead to a different operational benchmark. In order to avoid heterogeneity, some studies have used the route-based performance evaluation to substitute for the company-based performance evaluation (Chiou and Chen 2006; Lin et al. 2010; Yu and Chen 2011; Chiou et al. 2012).

2.4.2 Environmental Factors

Since firms run in different environments, their operation outcome will be affected by the environmental factors that they face. If environmental factors are ignored, performance measures would be seriously biased against firms that generate a misleading performance evaluation profile. For example, the population at the airport would affect its outputs. Higher utilization of an airport does not guarantee more efficient management, since some of the effects may be caused by higher population around the airport. It is appropriate to adjust for environmental conditions before credible results could be presented. Although, environmental factors usually cannot be controlled by the administrator, they may influence how we measure efficiency in the use of capacity. Standard DEA assumes that the assessed units are operated in similar operational environments (Golany and Roll 1989). Often the assumption of homogeneous environments is violated. Hence, it is essential that, if the model is to be used in this manner, factors which establish the operational environments need to be incorporated into the model. A number of different approaches have been developed to overcome this weakness (Syrjanen 2004). In this section, the approach introduced by Banker and Morey (1986) is described.

According to Banker and Morey (1986), a DMU should be compared with its peers under a similar operational environment. In order to capture the effects of environmental factors on the production and consumption process, we include the environmental variables as non-discretionary inputs by adding the following constraints into the network DEA model illustrated in Sect. 2.3:

$$\sum_{j=1}^{J} z_j^P e_{fj}^P \le e_{fk}^P, \ \ f = 1, \ldots, F, \tag{2.6.7}$$

$$\sum_{j=1}^{J} z_j^C e_{gj}^C \le e_{gk}^C, \ \ g = 1, \ldots, G, \tag{2.6.8}$$

where $e^P \in R_+^F$ and $e^C \in R_+^G$ represent environmental factors f and g associated only with the production and consumption processes of firm j, respectively.

2.4.3 Undesirable Outputs

Since undesirable outputs are often produced together with desirable outputs, the more complete performance evaluation of a transportation organization should consider the trade-off between the utilization of desirable output and the control of undesirable output. For example, aircraft noise has the greatest influence on the community surrounding the airport (Morrell and Lu 2000). If the effect of aircraft noise is ignored, the rank of airport performance in capacity utilization may be severely distorted. Thus, when the efficiency of airports is evaluated, the provision of desirable outputs like the number of passengers should be credited, but the provision of undesirable outputs like noise pollution should be penalized.

Following Färe et al. (1989) and Chung et al. (1997), we use a directional distance function to construct the efficiency measurement model that simultaneously credits a decrease in undesirable outputs and an increase in desirable outputs. Let $u^C \in R_+^H$ denote an undesirable output vector in the consumption process. Since, in the consumption process, DMUs seek to increase the desirable outputs and decrease the undesirable outputs simultaneously, the objective function of the network model still is (2.6). However, in the consumption process, an additional constraint must be added to present the deflation of undesirable outputs. This constraint is written as the form:

$$\sum_{j=1}^{J} z_j^C u_{hj}^C = \left(1 - \beta_k^C\right) u_{hk}^C, \quad h = 1, \ldots, H, \tag{2.6.9}$$

By applying the objective function identified in (2.6) and the constraints identified in Equations (2.6.1)–(2.6.9), we could compute the efficiency of transportation organizations based on the network structure with these undesirable outputs.

2.4.4 Multi-activity Framework

In many instances, organizations of any complexity typically consist of a number of individually identifiable units (Beasley 2003). For example, within a bus transit firm/railway company these units may correspond to various transportation services. Bus transit firms/railway companies may operate both highway and urban bus services/passenger and freight transportation services, what is efficient in a highway bus service/passenger transportation service may not be efficient in an urban bus service/freight transportation service, and thus different efficiency ratings for various activities should be distinguished. Units are linked by allocating resources, such as management labor and mechanics, to individual activities. The total amount of resources that the firm can allocate will be limited and unseparated. To allocate those unseparated shared resources is plainly important in a number of firms. However, the conventional DEA model evaluates the efficiency that a DMU transforms inputs into outputs. It assumes that a DMU is equally efficient in all its activities. Hence, the problem of a firm's efficiency which faces different production functions using shared inputs needs to be solved.

Many studies have been engaged to deal with this shared input problem in a practical organizational standpoint and a cost perspective (Golany et al. 1993; Golany and Tamir 1995; Beasley 1995, 2003; Mar Molinero 1996; Thanassoulis 1996; Färe et al. 1997, 2002; Mar Molinero and Tsai 1997; Tsai and Mar Molinero 1998, 2002; Cook and Kress 1999; Cook et al. 2000). The multi-activity DEA model, a novel refinement of the conventional DEA approaches, for the joint determination of efficiencies in the DEA context, was proposed by Beasley (1995) and subsequently revised by Mar Molinero (1996) and Tsai and Mar Molinero (1998, 2002). Specifically, the multi-activity model is used to evaluate efficiencies of organizations that engage in several activities simultaneously and some inputs and outputs are utilized and produced among all the activities.

In order to capture characteristics of the multi-activity model based on the network structure, we construct a multi-activity network DEA model by taking the railway companies, which generally provide passenger and freight transportation services in the production process, as example. A schematic of the performance evaluation in multi-activity network structure for a particular railway company is depicted in Fig. 2.3. In Fig. 2.3, the production process is divided into two sub-processes by passenger and freight transportation activities and those shared inputs are allocated to these two sub-processes.

Similarly, suppose there are J railway companies to be evaluated. We denote that $x^{PP} \in R_+^I$ and $m^{(PP,C)} \in R_+^N$ are (dedicated) inputs and intermediate outputs

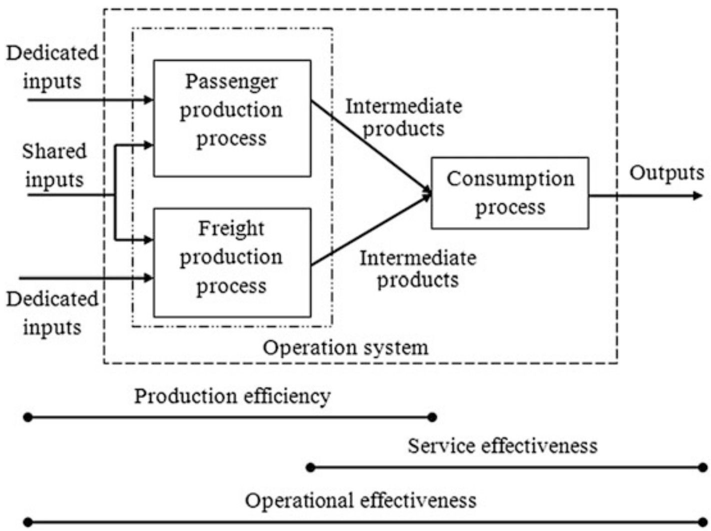


Fig. 2.3 Performance evaluation in the multi-activity network structure

associated solely with the passenger production process (PP), $x^{FP} \in R_{+}^{L}$ and $m^{(PP,\,C)} \in R_{+}^{N}$ are (dedicated) inputs and intermediate outputs associated solely with the freight production process (FP), but $x^{PFS} \in R_{+}^{M}$ are shared inputs associated in part with PP and in part with FP. Railway companies use (dedicated and shared) inputs to produce intermediate outputs in the production process. The intermediate products are consumed in consumption processes to produce final outputs, $y^{C} \in R_{+}^{D}$. In the situation where there are inputs associated with both activities, we assume that these shared inputs can be apportioned between PP and FP. In this way, each joint input contributes to the determination of the passenger efficiency and the freight efficiency in the production process. Assuming that the proportions of the shared inputs assigned to each one of the said activities are α_{PP} and $1 - \alpha_{PP}$. Thus the objective function of the multi-activity network DEA model is revised as follows:

$$Max\ \beta^{k} = w_{k}^{PP}\beta_{k}^{PP} + w_{k}^{FP}\beta_{k}^{FP} + w_{k}^{C}\beta_{k}^{C}, \tag{2.7}$$

where β_{k}^{PP} and β_{k}^{FP} measure the maximum deflation of inputs in the passenger and freight production processes, respectively; β_{k}^{C} measure the maximum inflation of outputs in the consumption processes; w_{k}^{PP}, w_{k}^{FP} and w_{k}^{C} are positive numbers which represent the relative importance of these activities/processes respectively, and $w_{k}^{PP} + w_{k}^{FP} + w_{k}^{C} = 1$. Equation 2.7 is subject to the following constraints:

The passenger production process is given by

$$\sum_{j=1}^{J} z_{j}^{PP} x_{ij}^{PP} \leq \left(1 - \beta_{k}^{PP}\right) x_{ik}^{PP},\ \ i = 1, \ldots,\ I, \tag{2.7.1}$$

$$\sum_{j=1}^{J} z_{j}^{PP} m_{nj}^{(PP,\,C)} \geq m_{nk}^{(PP,\,C)},\ \ n = 1, \ldots,\ N, \tag{2.7.2}$$

$$\beta_{k}^{PP} \geq 0,\ \ z_{j}^{PP} \geq 0,\ \ j = 1, \ldots, J. \tag{2.7.3}$$

The freight production process is given by

$$\sum_{j=1}^{J} z_{j}^{FP} x_{lj}^{FP} \leq \left(1 - \beta_{k}^{FP}\right) x_{lk}^{FP},\ \ l = 1, \ldots,\ L, \tag{2.7.4}$$

$$\sum_{j=1}^{J} z_{j}^{FP} m_{qj}^{(FP,\,C)} \geq m_{qk}^{(FP,\,C)},\ \ q = 1, \ldots,\ Q, \tag{2.7.5}$$

$$\beta_{k}^{FP} \geq 0,\ \ z_{j}^{FP} \geq 0,\ \ j = 1, \ldots, J. \tag{2.7.6}$$

The consumption process consists of

$$\sum_{j=1}^{J} z_j^C m_{nj}^{(PP,\,C)} \leq m_{nk}^{(PP,\,C)}, \;\; n = 1, \ldots, \;\; N, \tag{2.7.7}$$

$$\sum_{j=1}^{J} z_j^C m_{qj}^{(FP,\,C)} \leq m_{qk}^{(FP,\,C)}, \;\; q = 1, \ldots, \;\; Q, \tag{2.7.8}$$

$$\sum_{j=1}^{J} z_j^C y_{dj}^C \geq \left(1 + \beta_k^C\right) y_{dk}^C, \;\; d = 1, \ldots, \;\; D, \tag{2.7.9}$$

$$\beta_k^C \geq 0, \;\; z_j^C \geq 0, \;\; j = 1, \ldots, J, \tag{2.7.10}$$

Equations 2.7.11 and 2.7.12 represent the allocation of shared inputs to the passenger and freight production processes:

$$\sum_{j=1}^{J} \alpha_{PP} z_j^{PP} x_{mj}^{PFS} \leq \left(1 - \beta_k^{PP}\right) \alpha_{PP} x_{mk}^{PFS}, \;\; m = 1, \ldots, M \tag{2.7.11}$$

$$\sum_{j=1}^{J} (1 - \alpha_{PP}) z_j^{FP} x_{mj}^{PFS} \leq \left(1 - \beta_k^{FP}\right)(1 - \alpha_{PP}) x_{mk}^{PFS}, \;\; m = 1, \ldots, M \tag{2.7.12}$$

where z_j^{PP}, z_j^{FP} and z_j^C represent intensity variables for passenger production, freight production and consumption processes of each DMU j, respectively.

The objective function in (2.7) takes a value of zero if and only if the railway company's PP is technically efficient, its FP is technically efficient, and its consumption process is simultaneously serviced effectively. However, its value is greater than zero if and only if the railway company is technically inefficient at least one of the two sub-processes or the service is ineffective.

2.5 Examples

In this section, we provide related three cases to illustrate applications in empirical studies. First, a route-based performance evaluation in a network DEA model will be described. Next, a case that incorporates environmental factors and multiple activities into a network DEA model will be explored. Finally, we will investigate a multi-activity DEA model with these undesirable outputs.

2.5.1 Route-Based Network DEA Model[1]

To explore a route-based performance evaluation in a network DEA model, an example of 15 domestic air routes operated by a Taiwanese domestic airline in 2001 is applied. The performance of an air routes also can be divided into production efficiency (PE), service effectiveness (SE) and operational effectiveness (OE).

2.5.1.1 The Data

The input–output framework on the network model is depicted in Fig. 2.4. Input–output variables of an air routes are illustrated as follows:

1. *Output:* Number of passenger-miles.
2. *Inputs:* Personnel cost, fuel cost and aircraft cost.
3. *Intermediate output:* Number of seat-miles.

2.5.1.2 Empirical Results

Table 2.1 gives us a clear and complete picture of relative performance for the sample's air routes in three performance dimensions. It follows that for an air route to be able to locate on the overall operational effectiveness frontier, it needs to achieve both full production efficiency and service effectiveness. Hence, it can be found that there is a possibility of improvement for all air routes since their operational effectiveness scores are all less than unity. Table 2.1 also indicates that the average air routes' production efficiency, service effectiveness and operational effectiveness are 0.829, 0.833 and 0.689, with a standard deviation of 0.139, 0.099 and 0.135, respectively. It is worth mentioning that the scores of production efficiency and service effectiveness must be used together to identify which processes need to be improved. For example, the operational effectiveness of air routes TSA-KHH and TSA-MZG are about the same (their scores are 0.740 and 0.739,

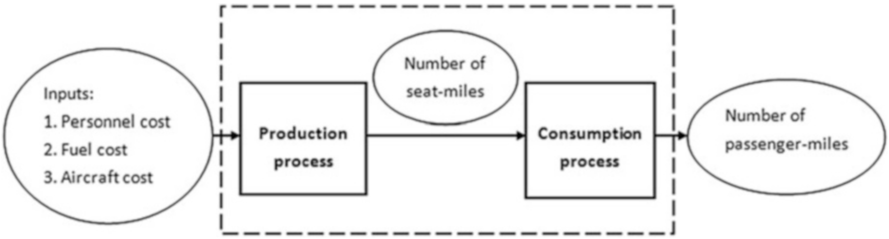


Fig. 2.4 Input–output variables in a network model

[1] Adapted from Yu and Chen (2011).

Table 2.1 Efficiency and effectiveness scores of the network model

	PE	SE	OE	Length	Aircraft	Seats	Market	Service area
TSA-KHH	1.000 (1)	0.740 (13)	0.740 (6)	183	MD-90	155	Business	Inland
TSA-TNN	0.975 (3)	0.645 (15)	0.629 (8)	164	MD-90	155	Business	Inland
TSA-TXG	0.625 (14)	0.807 (10)	0.504 (14)	77	DH8-300	56	Business	Inland
TSA-CYI	0.780 (10)	0.780 (11)	0.609 (11)	128	DH8-300	56	Recreation	Inland
TSA-TTT	0.895 (7)	0.684 (14)	0.612 (10)	161	MD-90/ DH8-300	155/ 56	Recreation	Inland
TSA-MZG	0.908 (6)	0.814 (9)	0.739 (7)	156	MD-90/ DH8-300	155/ 56	Recreation	Offshore
TXG-MZG	0.778 (11)	0.775 (12)	0.603 (12)	82	DH8-300	56	Recreation	Offshore
CYI-MZG	0.588 (15)	0.830 (8)	0.488 (15)	52	MD-90/ DH8-300	155/ 56	Recreation	Offshore
TNN-MZG	0.636 (13)	0.905 (5)	0.576 (13)	56	MD-90/ DH8-300	155/ 56	Recreation	Offshore
KHH-MZG	0.696 (12)	0.881 (6)	0.614 (9)	85	MD-90/ DH8-300	155/ 56	Recreation	Offshore
TSA-KNH	0.910 (5)	0.938 (2)	0.853 (3)	196	MD-90/ DH8-300	155/ 56	Recreation	Offshore
TXG-KNH	0.819 (9)	1.000 (1)	0.819 (4)	146	MD-90/ DH8-300	155/ 56	Recreation	Offshore
CYI-KNH	0.894 (8)	0.845 (7)	0.755 (5)	145	DH8-300	56	Recreation	Offshore
TNN-KNH	0.932 (4)	0.933 (3)	0.869 (2)	155	DH8-300	56	Recreation	Offshore
KHH-KNH	1.000 (1)	0.922 (4)	0.922 (1)	183	MD-90/ DH8-300	155/ 56	Recreation	Offshore
Max	1.000	1.000	0.922					
Min	0.588	0.645	0.488					
Mean	0.829	0.833	0.689					
SD	0.139	0.099	0.135					

Notes: Resources of the attributes of each air route are from Chiou and Chen (2006)

respectively). However, the activities they need to improve to achieve operational effectiveness frontier are different. Air route TSA-KHH, with production efficiency score = 1.000 and service effectiveness score = 0.740, only needs to expand its consumed output 35.1 % (1/0.740) to the service effectiveness frontier and then it will achieve operational effectiveness frontier. On the other hand, air route TSA-MZG, with production efficiency score = 0.908 and service effectiveness score = 0.814, needs to contract its input 9.2 % (1 − 0.908) and expand its consumed output 22.8 % (1/0.814) simultaneously to achieve operational effectiveness.

This is a possible indication that inferior production efficiency and/or service effectiveness cause operational ineffectiveness on air routes.

Next to the column of operational effectiveness score, Table 2.1 also shows some of the operational information including route length, aircraft type, operational market, and service areas for each air route. As indicated, there are five routes serving major inland cities and ten routes connecting these cities to two offshore cities. From the operational effectiveness point of view, offshore air routes are performing better on average in comparison to the inland air routes in the sample. In particular, the top five routes with higher operational effectiveness scores all belong to the offshore air routes. However, we should stress that the better performance of offshore air routes than inland air routes might not mainly come from the better management of the decision makers of those routes, but may be the result of limited substitution in transportation modes and the increasing demand from the tourism market offshore.

As for route length, long air routes perform better than short ones. This is intuitive, since one can easily realize that the longer the route is, the higher performance will be. First, bigger aircraft with more seats in general are used to serve longer distance travel. Secondly, shorter routes in general spend a longer proportion of their time in ground operations than long flights. The current results suggest that the sample airline needs to focus on improving performance of those short air routes. The above results show that the operational effectiveness of air routes is to a lesser extent due to the market types and to a greater extent due to the length and service area of air routes in the Taiwan domestic air transportation market.

Lastly, as it appears in Table 2.1, the use of different types of aircraft seems to show some effects on the air routes' service effectiveness but not production efficiency measure, since air routes operating with mixed types of aircraft appear to be more service effective than those using a single type of aircraft, while mixed type air routes do not perform better in production efficiency. A possible explanation is that a higher loading factor can be achieved if different types of aircraft are alternatively dispatched to serve peak demand (MD-90) and off-peak demand (DH8-300), while the benefits from lower operating cost does not guarantee better production efficiency. This implies that air routes operations need to meet the obligation of providing a fixed timetable of flights. This result recommends that the sample airline alternatively dispatch different types of aircraft to serve varying-demand routes to increase its air routes' service effectiveness.

2.5.2 Multi-activity DEA Model with Environmental Factors and Undesirable Output[2]

We provide an example for 24 Taiwan's multimode bus transit firms in 2001 that incorporate environmental factors (E) and undesirable output (U) into a multi-activity

[2] Adapted from Yu and Fan (2006).

DEA (MDEA) model to analyze the highway bus effectiveness (HBE), urban bus effectiveness (UBE) and operational effectiveness (OE) of each bus transit firm. All these firms operated both highway bus service (HB) and urban bus service (UB).

2.5.2.1 The Data

The input–output framework on the multi-activity model is portrayed in Fig. 2.5. Input–output variables of individual activities of a bus transit firm are illustrated as follows:

1. *Dedicated inputs of highway bus service:* Drivers, vehicles, fuel and network length in the highway bus sector.
2. *Desirable output of highway bus service:* Passenger-km.
3. *Dedicated inputs of urban bus service:* Drivers, vehicles, fuel and network length in the urban bus sector.
4. *Desirable output of urban bus service:* Passengers.

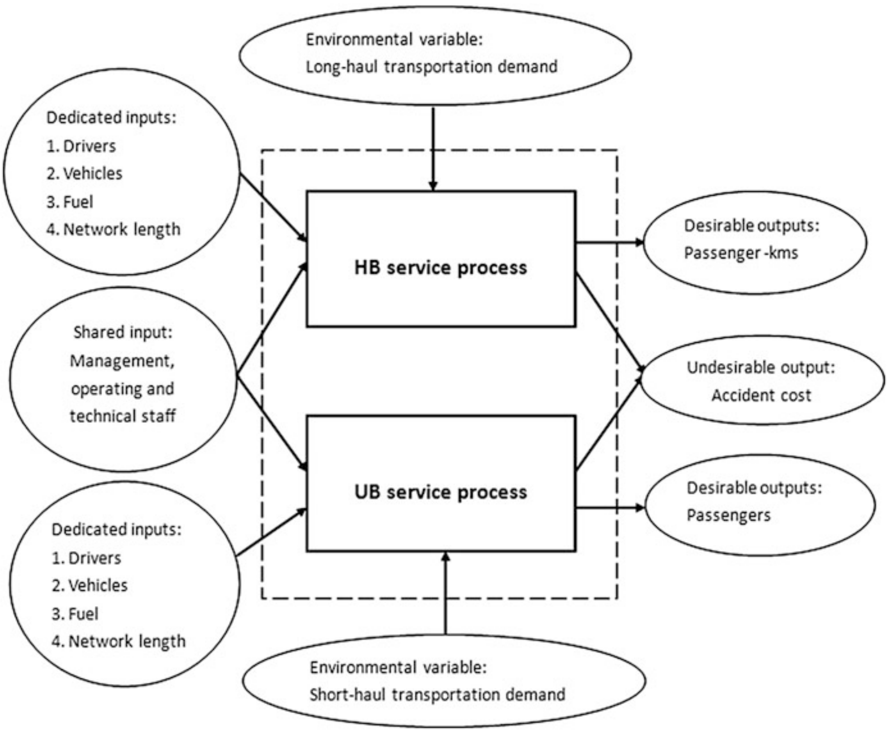


Fig. 2.5 Input–output variables in a multi-activity model

5. *Shared input for highway and urban bus services:* Management, operating and technical staff.
6. *Undesirable output for highway and urban bus services:* Accident cost.
7. *Environmental variable of highway bus service:* Long-haul transportation demand.
8. *Environmental variable of urban bus service:* short-haul transportation demand.

2.5.2.2 Empirical Results

For each firm, four overall operational effectiveness measures have been calculated by different DEA models, as shown in Table 2.2. Note that all the operational effectiveness scores should be less than or equal to unity and that a higher score indicates a more effective status. The first column is the overall operational effectiveness obtained from a conventional DEA model. These conventional indices diverge from 0.523 to 1.0 with a mean level of 0.952. The number and percentage of the fully operationally effective units is 17 and 70.83 % of the 24 bus firms. As the second column indicates, the overall operational effectiveness indices obtained from the multi-activity DEA model 1 have larger mean value ranges, from 0.421 to 1.0, with a mean overall operational effectiveness of 0.850. Moreover, only four out of 24 firms are operationally effective. Column 5 reports the overall operational effectiveness scores obtained from the multi-activity DEA model 2 which includes an environmental factor, but ignores undesirable output side effects. As can be noted, the estimated effectiveness diverges substantially from 0.570 to 1.0 with a mean value of 0.898. Of the 24 bus firms analyzed, only five are deemed effective. The results of column 8 are obtained from the multi-activity DEA model 3 in which the overall operational effectiveness of a firm is evaluated on the basis of its ability to increase desirable outputs and reduce inputs and undesirable output simultaneously. The overall operational effectiveness scores vary from 0.576 to 1.0 with a mean effectiveness score of 0.884. The number and percentage of the fully operationally effective units increases to 7 and 29.17 % of the 24 bus firms as the undesirable output is included. If we concentrate on the highway bus service, ten of the bus firms exhibit operationally effective behavior that is superior to the rest. With regards to urban transit, a maximum level of effectiveness is achieved by nine firms, with bus firms that are operationally effective in each of the two services coinciding in only seven cases.

These above results imply that the conventional DEA operational effectiveness measure may be seriously misleading if it ignores the operational effectiveness of firms, which carry out various activities whilst sharing common resources. In addition, for those bus firms where environmental factors and undesirable output are important, the illustration shows that different multi-activity DEA models lead to different results. The multi-activity DEA model 3 provides a deep structure that more fully takes the shared inputs, environmental factors and undesirable output into consideration.

Table 2.2 Conventional DEA and multi-activity DEA effectiveness measures under different situation

	Conventional (DEA)	MDEA model 1 (*excluding E and U*)			MDEA model 2 (*including E only*)			MDEA model 3 (*including E and U*)		
	OE	OE	HBE	UBE	OE	HBE	UBE	OE	HBE	UBE
Max	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Min	0.523	0.421	0.405	0.167	0.570	0.708	0.432	0.576	0.559	0.310
Mean	0.952	0.850	0.846	0.854	0.898	0.903	0.893	0.884	0.912	0.856
SD	0.118	0.171	0.160	0.230	0.120	0.105	0.161	0.135	0.117	0.218
No. of fully effective units	17	4	7	7	5	8	8	7	10	9
% of fully effective units	70.83	16.67	29.17	29.17	20.83	33.33	33.33	29.17	41.67	37.50

2.5.3 Multi-activity Network DEA Model with Environmental Factors[3]

Furthermore, the operational process of a multimode bus transit firms can be divided into two sub-processes: production and consumption processes. In addition, the production process includes two activities: highway bus service (HB) and urban bus service (UB). Hence, we also apply this example for multimode bus transit firms to illustrate the performance obtained from multi-activity network DEA model, but, in this section, the used model incorporates multiple activities, multiple processes and environmental factors to analyze the highway bus efficiency, urban bus efficiency, production efficiency, service effectiveness and operational effectiveness of each bus transit firm. The data set used in the measurement of performance in Taiwan's bus transit system comprised a sample of 23 firms located all over the island in 2001 and 2002. All these firms operated both highway bus service and urban bus service.

2.5.3.1 The Data

The input–output framework on the multi-activity network model is represented in Fig. 2.6. Input–output variables and environmental variables of individual activities and processes of a bus transit firm are illustrated as follows:

1. *Dedicated inputs of highway bus production service:* Drivers, vehicles, fuel and network length in the highway bus sector.
2. *Intermediate output of highway bus production service:* Vehicle-kms in the highway bus sector.
3. *Dedicated inputs of urban bus production service:* Drivers, vehicles, fuel and network length in the urban bus sector.
4. *Intermediate output of urban bus production service:* Vehicle-kms in the urban bus sector.
5. *Dedicated input in consumption process:* Sales staff.
6. *Output in consumption process:* Passenger-kms and passengers.[4]
7. *Shared input for highway and urban bus production services:* Mechanics.
8. *Shared input for highway bus production service, urban bus production service and consumption process:* Management employees.
9. *Environmental variables:* Population density and car ownership.

[3] Adapted From Yu and Fan (2009).

[4] The passenger-km are not available for UB service, so the number of passengers is used as a proxy variable in this paper. It is more appropriate to use passenger-kms as final output variables.

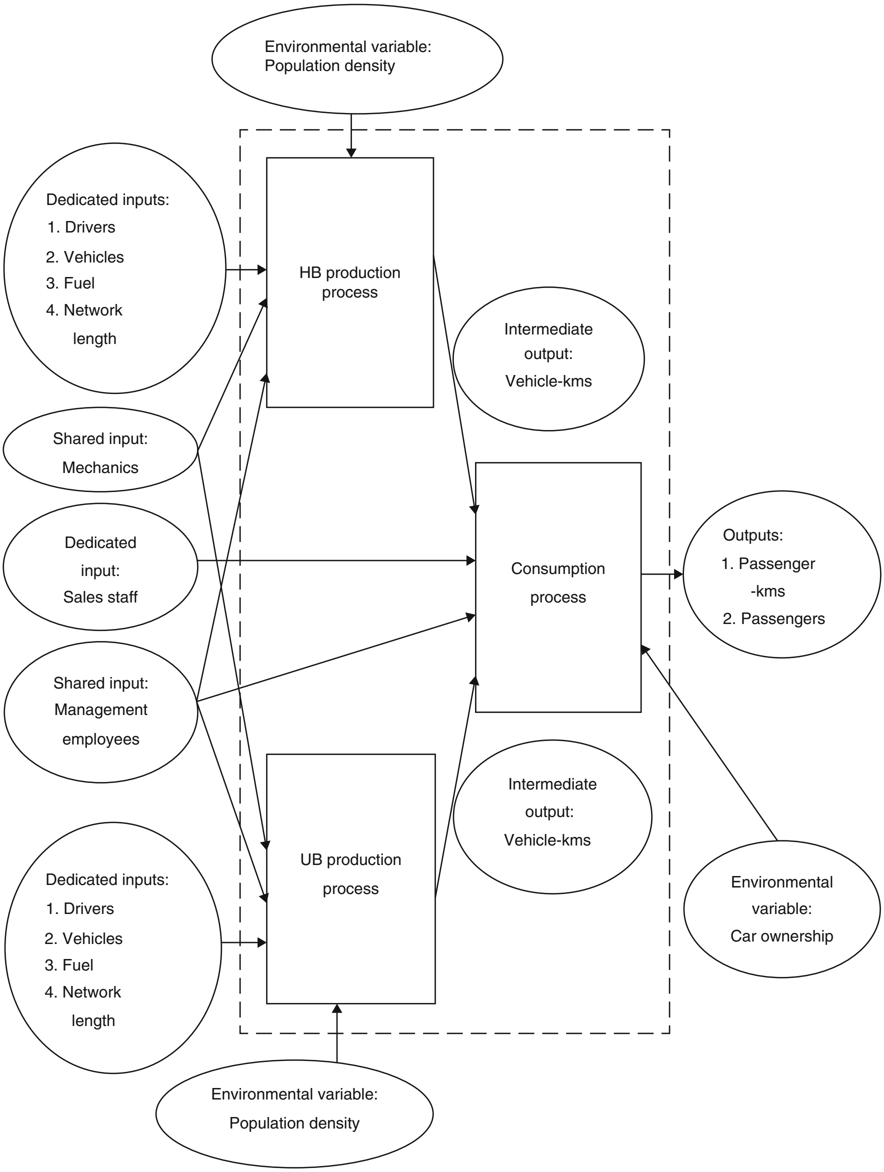


Fig. 2.6 Input–output variables in a multi-activity network model

2.5.3.2 Empirical Results

In this section, we present estimates of performance measures based on the all-in-one multi-activity network DEA model, and three separate conventional DEA models. It is worth noting that the production efficiency, service effectiveness and

Table 2.3 Efficiency and effectiveness scores of the multi-activity network model

	Highway bus efficiency $(1-\beta_k^H)$	Urban bus efficiency $(1-\beta_k^U)$	Production efficiency (β_k^P)	Service effectiveness $(1+\beta_k^C)$	Operational effectiveness $(1+\beta_K)$
Max	1.000	1.000	1.000	2.502	1.837
Min	0.613	0.514	0.738	1.000	1.000
Mean	0.894	0.864	0.879	1.160	1.141
SD	0.103	0.141	0.073	0.329	0.171

Notes: (1) Each of the efficiency or effectiveness scores is the mean of the estimated values of 2 years' observations; (2) $\beta_k^P = 1 - \left(\frac{\beta_k^H + \beta_k^U}{2}\right)$; and (3) $\beta_k = 0.25\beta_k^H + 0.25\beta_k^U + 0.5\beta_k^C$

operational effectiveness estimated by the multi-activity network DEA model imply that those performance measures are not independent. The results of multi-activity network DEA are summarized in Table 2.3. If the value of the production efficiency is equal to unity, this denotes that it is "efficient", whereas values less than 1 indicate that it is "inefficient". On the other hand, if the value of the service effectiveness or operational effectiveness is equal to unity, this denotes that it is "effective", whereas values greater than 1 denote that it is "ineffective".

In the first two columns, the highway bus efficiency and urban bus efficiency, and in the fourth column, the service effectiveness, are evaluated on the basis of their ability to share common inputs among different activities, and to determine simultaneously their efficiency and effectiveness. With regard to the average production efficiency, the means of highway and urban bus efficiencies are lower than 1, indicating that there was inefficient in the production process for the sample as a whole. When the mean of service effectiveness score is greater than 1, in this case 1.160, this denotes an "ineffective" score for the sample as a whole. This service effectiveness may be explained by the inability of firms to expand ridership, as the vehicle-km provision cannot be reduced under the same environment. The average operational effectiveness was also greater than 1 (1.141), indicating that the sample as a whole was "ineffective". For efficient firms that are efficient in regard to their production but not consumption processes, it is implied that they operate ineffectively, and hence there is further improvement in terms of service effectiveness. The managers could pay more attention to increasing the utilization of the produced service to improve their service effectiveness. For firms that are inefficient in their production processes but effective in their consumption processes, it implies that they are not production efficient. This could mean that firms should reduce their input proportions with respect to their frontiers in order to determine the improvement needed in each activity to catch up with the frontier firms.

Based on the comparison, efficiency and effectiveness measurements are examined, and are depicted in Table 2.4. The production efficiency index in the multi-activity network model has slightly lower efficiency score, and only 3 of the 23 firms are operating on the production frontier, while 9 of the 23 are operating efficiently on the production frontier under the conventional model. With respect to service effectiveness, the results reveal a relatively lower effectiveness score (lower

Table 2.4 Descriptive statistics of the conventional and multi-activity network models' performance scores and the results of the test of significance

	All samples	Production efficiency	Service effectiveness	Operational effectiveness
Multi-activity network model				
Number of firms	69	23	23	23
Number of efficient or effective scores	12	3	8	1
Number of inefficient or ineffective scores	57	20	15	22
Mean of efficiency or effectiveness scores	–	0.879	1.160	1.141
Conventional model				
Number of firms	69	23	23	23
Number of efficient or effective scores	21	9	6	6
Number of inefficient or ineffective scores	48	14	17	17
Mean of efficiency or effectiveness scores	–	0.965	1.237	1.144
Correlations				
Network vs. conventional	0.901	0.471	0.935	0.858
Test of significance				
p-value	0.003**	0.000**	0.097*	0.885

Notes: "*" and "**" mean significant at the 10 % and 5 % level of significance, respectively

effectiveness score represents more effective) than the conventional DEA model. As to operational effectiveness, the results also indicate that the average effectiveness score is relatively lower (representing more effective).

In order to provide statistically robust findings about these transit firms' respective performances, paired difference experiments are applied. This experiment is conducted to verify whether the sample firms for the two kinds of models were drawn from the same performance populations for the three measures, respectively. The significance of paired comparisons is that it is based on a two-tailed test at the 0.05 acceptance level. As shown in Table 2.4, the test of significance yielded a p-value of 0.000 of production efficiency, which shows a statistically significant difference in terms of production efficiency. However, the statistical test confirmed that the service effectiveness and operational effectiveness measures were not significantly different, having p-values of 0.097 and 0.885, respectively. On the other hand, the statistical test for the entire sample, which pooled the three measures in a set, yielded a p-value of 0.003 which reveals a significant difference between the two models at the 5 % acceptance level. The results of the statistical tests for the two models may imply that the significant difference in production efficiency estimated by the mixed structure network and conventional models gave rise to the significant differences in the overall samples for these three measures, even though the differences between the service effectiveness and operational effectiveness measures estimated by these two models are generally insignificant. Therefore some more means are applied for further comparison.

The results obtained from the multi-activity network and the conventional models are quite different in terms of efficient or effective units. In general, the multi-activity network model is more demanding than the conventional one. This is explained by the following two facts. First, the achievement of a better degree of efficiency or effectiveness in the multi-activity network model requires that good productive and consumption matching behaviors are demonstrated on the part of the two services (HB and UB) as well as between the production and consumption processes, respectively. However, with the conventional model, it is possible that there are compensations between the two production activities and one consumption process in such a way that one firm will always achieve the production frontier provided that, in global terms, it demonstrates behavior which is superior to the rest, even if such superiority is not demonstrated in all the activities (services) it carries out. Second, a representation of both production and consumption processes in a unified framework is allowed in the multi-activity network model, and hence the three measures interact to determine the performance, while with the conventional model the three measures are calculated independently, even though there is a high degree of correlation between individual scores (service and operational effectiveness) obtained from the multi-activity network DEA model and those derived from the conventional DEA model. This indicates that the multi-activity network DEA model provides a nearly coincident result in terms of service and operational effectiveness, while it is worth noting that production efficiency is quite different. It is more reasonable to use the results of the multi-activity network DEA model for gauging the transit firms' performance, since the potential benefit of this model is that it provides the possibility of looking deeply into the production and consumption processes. This shows that by considering the multiple activities and unstorable characteristics of transit services in the network model, firms may not only compare their performances with those of peer groups under practical and realistic conditions, but the inter-related effects caused by the various activities and processes may also be considered.

2.6 Conclusions

In this chapter, we describe a network graph of operational structure in the transportation sector to represent the operational characteristics of transportation services, and apply this concept to construct a network DEA model that illustrates the operational behavior in the sense of maximization of consumed outputs and minimization of initial inputs. To document its practicality, the network DEA model provides a deeper structure that takes unstorable characteristics of transportation services into consideration. Since the focus of the chapter is on providing a more reasonable performance measurement in the transportation sector and how the DEA model can be applied practically, we further incorporate route-based performance evaluation, environmental factors, undesirable outputs and multi-activity framework into the network DEA model, respectively. These models can provide the sources of inefficiency within a transportation organization. Identification of such sources can help managers to design the implementation of operational policies and management strategies to improve performance. In addition, we have provided three relative applications in transportation organizations to illustrate the selection of inputs and outputs as well as the results.

References

Adler N, Golany B (2001) Evaluation of deregulated airline networks using data envelopment analysis combined with principal component analysis with an application to Western Europe. Eur J Oper Res 132:260–273

Banker RD, Morey R (1986) Efficiency analysis for exogenously fixed inputs and outputs. Oper Res 34:513–521

Beasley JE (1995) Determining teaching and research efficiencies. J Oper Res Soc 46:441–452

Beasley JE (2003) Allocating fixed costs and resources via data envelopment analysis. Eur J Oper Res 147:198–216

Boame AK (2004) The technical efficiency of Canadian urban transit systems. Transp Res Part E 40:401–416

Borger BD, Kerstens K, Costa A (2002) Public transit performance: what does one learn from frontier studies? Transp Rev 22(1):1–38

Chiou YC, Chen YH (2006) Route-based performance evaluation of Taiwanese domestic airlines using data envelopment analysis. Transp Res E 42:116–127

Chiou YC, Lan LW, Yen BTH (2012) Route-based data envelopment analysis models. Transp Res E 48:415–425

Chu X, Fielding GJ, Lamar BW (1992) Measuring transit performance using data envelopment analysis. Transp Res A 26(3):223–230

Chung YH, Färe R, Grosskopf S (1997) Productivity and undesirable outputs: a directional distance function approach. J Environ Manage 51:229–240

Cook WD, Hababou M, Tuenter HJH (2000) Multicomponent efficiency measurement and shared inputs in data envelopment analysis: an application to sales and service performance in bank branches. J Prod Anal 14:209–224

Cook WD, Kress M (1999) Characterizing an equitable allocation of shared costs: a DEA approach. Eur J Oper Res 199:652–661

Cowie J (2002) Acquistion, efficiency and scale economies: an analysis of the British bus industry. Transport Rev 22(2):147–157

Cowie J, Asenova D (1999) Organization form, scale effects and efficiency in British bus industry. Transportation 26:231–248

Färe R, Grosskopf S, Lovell CAK, Pasurka C (1989) Multilateral productivity comparisons when some outputs are undesirable: a nonparametric approach. Rev Econ Stat 71:90–98

Färe R, Grosskopf S (1996) Productivity and intermediate products: a frontier approach. Econ Lett 50:65–70

Färe R, Grosskopf S (2000) Network DEA. Socio Econ Plan Sci 34:35–49

Färe R, Grosskopf S, Weber W (1997) The effect of risk-based capital requirement on profit efficiency in banking. Discussion Paper Series, Department of Economics, Southern Illinois University at Carbondale, pp 97–112

Färe R, Grosskopf S, Zaim O (2002) Hyperbolic efficiency and return to the dollar. Eur J Oper Res 131:671–679

Fielding GJ, Babitsky TT, Brenner ME (1985) Performance evaluation for bus transit. Transp Res 19(1):73–82

Fielding G (1992) Transit performance evaluation in the U.S.A. Transp Res A 26(6):483–491

Graham DJ (2008) Productivity and efficiency in urban railways: parametric and non-parametric estimates. Transp Res E 44:84–99

Golany B, Phillips FY, Rousseau JJ (1993) Models for improved effectiveness based on DEA efficiency results. IIE Trans 25:2–10

Golany B, Roll Y (1989) An application procedure for DEA. OMEGA Int J Manage Sci 17:237–250

Golany B, Tamir E (1995) Evaluation efficiency-effectiveness-equality trade-offs: a data envelopment analysis approach. Manage Sci 41:1172–1184

Hensher AA (1992) Total factor productivity growth and endogenous demand: establishing a benchmark index for the selection of operational performance measures in public bus firms. Transp Res B 26(6):435–448

Karlaftis MG (2003) Investigating transit production and performance: a programming approach. Transp Res A 37:225–240

Karlaftis MG (2004) A DEA approach for evaluating the efficiency and effectiveness of urban transit systems. Eur J Oper Res 152(2):354–364

Kerstens K (1996) Technical efficiency measurement and explanation of French urban transit companies. Transp Res A 30(6):431–452

Lan LW, Lin ETJ (2003) Technical efficiency and service effectiveness for the railway industry: DEA approaches. J East Asia Soc Transp Stud 5:2932–2947

Lan LW, Lin ETJ (2005) Measuring railway performance with adjustment of environmental effects, data noise and slacks. Transportmetrica 1:161–189

Lee D (1989) Transit cost and performance measurement. Transp Res A 9:147–170

Lin ETJ, Lan LW, Hsu CST (2010) Assessing the on-road route efficiency for an air-express courier. J Adv Transport 44:256–266

Luenberger DG (1992) Benefit functions and duality. J Math Econ 21:461–481

Mackie P, Nash C (1982) Efficiency and performance indicators: the case of the bus industry. Public Money Manage 3:41–44

Mar Molinero C (1996) On the joint determination of efficiencies in a data envelopment analysis context. J Oper Res Soc 47:1273–1279

Mar Molinero C, Tsai PF (1997) Some mathematical properties of a DEA model for the joint determination of efficiencies. J Oper Res Soc 48:51–56

McMullen BS, Noh DW (2007) Accounting for emissions in the measurement of transit agency efficiency: a directional distance function approach. Transp Res D 12:1–9

Morrell P, Lu CHY (2000) Aircraft noise social cost and charge mechanisms—a case study of Amsterdam Airport Schiphol. Transp Res D 5:305–320

Nolan JF, Ritchie PC, Rowcroft JE (2001) Measuring efficiency in the public sector using nonparametric estimators: a study of transit agencies in the USA. Appl Econ 33:913–922

Nolan JF, Ritchie PC, Rowcroft JE (2002) Identifying and measuring public policy goals: ISTEA and the US bus transit industry. J Econ Behav Organ 48:291–304

Oben K (1994) The economic cost of subsidy-induced technical inefficiency. Int J Transport Econ 21(1):3–20

Sheth C, Triantis K, Teodorovic D (2007) Performance evaluation of bus routes: a provider and passenger perspective. Transp Res E 43:453–478

Syrjanen MJ (2004) Non-discrectionary and discrectionary factors and scale in data envelopment analysis. Eur J Oper Res 158:20–33

Thanassoulis E (1996) A data envelopment analysis approach to clustering operating units for resource allocation purposes. OMEGA Int J Manage Sci 24:463–476

Thanassoulis E, Boussofiane A, Dyson RG (1996) A comparison of data envelopment analysis and ratio analysis as tools for performance assessment. OMEGA Int J Manage Sci 24(3):229–244

Tofallis C (1997) Input efficiency profiling: an application to airlines. Comput Oper Res 24(3):253–258

Tomazinis AR (1975) Productivity, efficiency, and quality in urban transportation systems. Lexington Books, Lexington

Tsai PF, Mar Molinero C (1998) The joint determination of efficiencies in DEA: an application to the UK health service. Discussion Paper, Department of Management, University of Southampton

Tsai PF, Mar Molinero C (2002) A variable returns to scale data envelopment analysis model for the joint determination of efficiencies with an example of the UK health service. Eur J Oper Res 31:21–38

Tulkens H (1993) On FDH, efficiency analysis: some methodological issues and applications to retail banking, courts, and urban transit. J Prod Anal 4:183–210

Viton P (1998) Changes in multi-mode bus transit efficiency, 1988–1992. Transportation 2:1–21

Yu MM (2007) Productivity change and the effects of enhancement of mass transportation program on the bus transit system in Taiwan. Transport Rev 29(5):383–407

Yu MM, Chen PC (2011) Measuring air routes performance using a fractional network data envelopment analysis model. Cent Eur J Oper Res 19(1):81–98

Yu MM, Fan CK (2006) Measuring the cost effectiveness of multimode bus transit in the presence of accident risks. Transp Plan Techn 29(5):383–407

Yu MM, Fan CK (2009) Measuring the performance of multimode bus transit: a mixed structure network DEA model. Transp Res E 45:501–515

Chapter 3
Total-Factor Energy Efficiency and Its Extensions: Introduction, Computation and Application

Jin-Li Hu and Tzu-Pu Chang

Abstract This chapter demonstrates how to use different types of DEA models to compute the total-factor energy efficiency (TFEE) scores, including CCR, Russell, QFI, SBM, DF, and DDF models. The TFEE is a disaggregate input efficiency index. Moreover, the TFEE framework which uses cross-section data can be extended to the total-factor energy productivity (TFEP) growth index by following Malmquist, Leunberger, and Malmquist-Leunberger models which use panel data. Finally, the regional data of Chinese regions during 2010–2011 with inputs and desirable as well as undesirable outputs are used for illustrating the computation of TFEE and TFEP scores.

Keywords Input efficiency • Output efficiency • Radial adjustment • Slack-based measure (SBM) • Fixed inputs • Malmquist productivity index • Leunberger productivity index

3.1 Introduction

Energy efficiency, defined as economic/physical output divided by energy input, is a well-known indicator to realize how energy inputs are efficiently used. However, there is a mainly drawback of this conventional indicator, that is traditional energy efficiency indicator only takes account of energy as single input. In other words, this indicator may neglect the substitution or complement among energy and other inputs, such as labors and capital (Chang and Hu 2010). As a result, it may obtain a plausible result if we use the partial factor energy efficiency indicator.

J.-L. Hu (✉)
Institute of Business and Management, National Chiao Tung University, Hsinchu, Taiwan
e-mail: jinlihu@gmail.com

T.-P. Chang
Department of Finance, National Yunlin University of Science and Technology, Yunlin, Taiwan

J. Zhu (ed.), *Data Envelopment Analysis*, International Series in Operations Research & Management Science 238, DOI 10.1007/978-1-4899-7684-0_3

To deal with above-mentioned issue, several researches have attempted to establish advanced measures over the past decade. Hu and Wang (2006), one of famous literatures on this subject, first propose a total-factor energy efficiency (TFEE) to calculate energy efficiency under a total-factor framework. TFEE is a relative efficiency index of energy usage and easy to understand and extend. Basically, the nature of this measure is to compute a ratio of target (efficient) energy input to actual energy input based on multiple inputs and outputs.

The next question is, however, how to find out the amount of target/efficient energy input. In fact, data envelopment analysis (DEA), a non-parametric frontier-based method, must be an appropriate technique for this concern. As we know, DEA is not only a tool for investigating decision making units' (DMU) performance/efficiency, but a tool for finding target/efficient inputs and outputs. More specifically, DEA can calculate input and output slacks, which is the amounts deviate from efficient frontier. Following this method, the excess energy input is deemed a kind of input slack and target/efficient energy input can be measured by the difference between actual energy input and slack. Actually, TFEE is built and computed by DEA approach in Hu and Wang (2006). With similar concepts, a large number of studies use DEA to construct kinds of energy efficiency index (e.g., Shi et al. 2010; Wei et al. 2009; Zhou and Ang 2008).

The original DEA model, as initially proposed by Charnes et al. (1978), is built on the work of Farrell (1957). Over the past three decades, DEA has been rapidly developed and applied in many research fields, such as banking and economic issues. Of course, DEA is also well accepted as a major tool for energy efficiency analysis. For details, Zhou et al. (2008) provide an outstanding survey. Their paper reviews a total of 100 studies published from 1983 to 2006 and discusses several relevant technical issues of DEA in energy efficiency research.

In terms of our paper, we focus on how to apply DEA approach to calculate energy efficiency. Along with the progress of DEA methodology, several extensions of basic DEA models have been introduced under different restrictions and assumptions (e.g., Ouellette and Vierstraete 2004; Zhou et al. 2006). Therefore, this paper not only introduces some advanced DEA models in energy efficiency field, but discusses the link between TFEE and these models. Furthermore, we will present an application of China's energy efficiency issue to show how to obtain energy efficiency from each DEA model. It is believed that this paper can benefit researchers, policy makers and students who are interested in energy efficiency study by using DEA approach.

The remainder of this paper is organized as follows. Section 3.2 illustrates the concepts of total-factor input and output efficiency. Section 3.3 introduces several DEA models related to energy efficiency issue. Section 3.4 provides an application of how to compute provincial energy efficiency in China through incorporating different kinds of DEA models. Section 3.5 discusses some extended research issues and Sect. 3.6 concludes this paper.

3.2 The Concept of Total-Factor Input and Output Efficiency

As mentioned above, traditional energy efficiency indicator is a partial factor measurement that disregards other important input factors. Thus, a total-factor indicator of energy use is useful and needed. Hu and Wang (2006) introduce total-factor energy efficiency index and overcome possible bias in traditional energy efficiency indicator. There are third key features of this index: first, TFEE can cope with multiple inputs and outputs, meaning that it considers a total-factor framework. Second, TFEE rescales traditional energy efficiency to a number ranging from zero to unity based on the production frontier in each period. Third, target energy input can be detected that is more important for policy makers.

By definition, TFEE is a ratio of target energy input to actual energy input. Because actual energy input is always greater than or equal to target energy input, TFEE must be smaller than unity. That is,

$$0 \leq \frac{Target\ Energy\ Input}{Actual\ Energy\ Input} \leq 1. \tag{3.1}$$

In addition, excess energy input is denoted as the gap between target energy input and actual energy input is a kind of slack. Thus, total energy slack, ranging from zero to positive infinite, is equal to actual energy input minus target energy input. Then, (3.1) can be illustrated by

$$0 \leq 1 - \frac{Total\ Energy\ Slack}{Actual\ Energy\ Input} \leq 1. \tag{3.2}$$

It is undoubted that the idea of TFEE is very simple and easy to understand. Generally speaking, on the one hand, TFEE is a special case of total-factor input efficiency if we only emphasize the efficient energy input. In other words, one can focus on other type of input variable and calculate the total-factor input efficiency of what is interested in. For example, Hu et al. (2006) investigate the total-factor water efficiency in regions of China through the foregoing concept.

On the other hand, obviously, we can derive a total-factor desirable output efficiency index. It is noted that actual desirable output is always small than or equal to target desirable output. Hence, a reciprocal representation of (3.1) would range from zero to unity. Then, total-factor desirable output efficiency is defined as:

$$0 \leq \frac{Actual\ Desirable\ Output}{Target\ Desirable\ Output} \leq 1. \tag{3.3}$$

Accordingly, total-factor desirable output efficiency can also be represented by

$$0 \leq 1 - \frac{Total\ Desirable\ Output\ Slack}{Target\ Desirable\ Output} \leq 1, \tag{3.4}$$

because total desirable output slack is equal to target desirable output minus actual desirable output. However, in the real world, undesirable outputs are inevitably produced within economic activities, especially in energy consumption process. It means that as people use energy input to generate economic outputs or other kinds of desirable outputs, they also produce some types of undesirable outputs such as carbon dioxide and sulfur dioxide. Therefore, a total-factor undesirable output efficiency index should be established indeed.

It is not surprising that total-factor undesirable output efficiency can be constructed following aforementioned idea.[1] The fact that the concept of total-factor undesirable output efficiency is similar to total-factor input efficiency because of minimization consideration. That is, actual undesirable output must be greater than or equal to target undesirable output and total undesirable output slack is greater than or equal to zero. Hence, total-factor undesirable output efficiency is represented

$$0 \le \frac{Target\ Undesirable\ Output}{Actual\ Undesirable\ Output} \le 1, \tag{3.5}$$

or

$$0 \le 1 - \frac{Total\ Undesirable\ Output\ Slack}{Actual\ Undesirable\ Output} \le 1. \tag{3.6}$$

Herein, we have briefly demonstrated the concept of total-factor input and output efficiency measures. Actually, total-factor input and output efficiency indices can explicitly tell us how much amounts of targets or abatements of interested inputs and outputs are. It is considered that these measures are very important and useful for policy makers and the governments. Moreover, the next step is to calculate target values or total slacks of inputs/outputs. This paper focuses on total-factor energy efficiency and introduces several basic and advanced DEA models to deal with this job in the following section.

3.3 DEA Models for Energy Efficiency Study

Since the works of Charnes et al. (1978) and Banker et al. (1984), there has experienced a vigorous expansion of DEA methodology. For instance, the efficiency measure of DEA models can be calculated by one of radial, non-radial, slack-based, and directional distance function models.[2] In the following subsections, therefore, present paper will introduce three groups of commonly used and well accepted DEA models related to energy efficiency issues.

---

[1] Zhou et al. (2010) construct a total-factor carbon emission performance measurement of which the concept is similar to our total-factor undesirable output efficiency.

[2] Zhou et al. (2008) present a clear structure of DEA models. According to their survey, a DEA model can be characterized by its reference technology and efficiency measure.

At first, let us illustrate the general notations that are used in all models. Assume that there are N DMUs that may involve countries, regions, sectors, etc. These DMUs consume energy and M non-energy inputs to produce R desirable outputs and K undesirable outputs.[3] Hence, the energy input and m-th non-energy input variable of the o-th DMU are represented by e_o and x_{mo}, respectively. Moreover, the r-th desirable output and k-th undesirable output variable of the o-th DMU are represented by y_{ro} and u_{ko}, respectively. Sequentially, we discuss the linkage between these models and the concept of total-factor energy efficiency as follows.

3.3.1 Proportional Adjustment Models Without Undesirable Outputs

The first category of energy-related DEA models is traditional proportional adjustment models without undesirable outputs. This paper introduces three kinds of models according to different assumption: the first one assumes all inputs can be radially adjusted, which is known as radial DEA model. The second one assumes all inputs can be proportionally adjusted with different ratios, which is known as non-radial DEA model. The third model assumes only energy can be adjusted and other inputs are quasi-fixed.

3.3.1.1 All Inputs Can Be Radially Adjusted

Based on the assumption of all inputs can be radially adjusted, there are two classical and widely used DEA models which are proposed by Charnes et al. (1978) and Banker et al. (1984), respectively. The former model is also called by CCR model and the latter is known as BCC model. The major feature of these models is that the adjustments of all inputs can be proportionally contracted without decreasing the amounts of current outputs. However, these two models have a distinct property on returns to scale (RTS), indicating that CCR model presents constant returns to scale (CRS), while BCC model exhibits variant returns to scale (VRS).

Hu and Wang's original TFEE is built and computed by an input-oriented CCR model. Accordingly, the linear programming (LP) problem for the o-th DMU can be solved by

[3] Most of energy efficiency studies using country-level data only take total energy consumption as energy input. However, several literatures use disaggregated energy inputs to analyze region- or sector-level data. For example, Honma and Hu (2008) investigate total-factor energy efficiency of 11 energy inputs of regions in Japan. For simplicity, we choose total energy consumption as only one energy input to illustrate energy-related DEA models. One can straightforward extend these models with multiple energy inputs indeed.

$$
\begin{aligned}
&\theta^* = \min \theta \\
&\text{s.t.} \quad \sum_{n=1}^{N} \lambda_n e_n \le \theta e_o, \\
&\sum_{n=1}^{N} \lambda_n x_{mn} \le \theta x_{mo}, \qquad m = 1, 2, \ldots, M, \\
&\sum_{n=1}^{N} \lambda_n y_{rn} \ge y_{ro}, \qquad r = 1, 2, \ldots, R, \\
&\lambda_n \ge 0, \quad n = 1, 2, \ldots, N.
\end{aligned} \tag{3.7}
$$

Equation 3.7 shows the input-oriented CCR model in envelopment form of which θ is a scalar and λ_n is an $N \times 1$ vector of constants.[4] The optimal solution, θ^*, yields an efficiency score for the o-th DMU. It is noted that, however, CCR model ignores the presence of non-radial slacks and results in a weak efficiency measure (Cooper et al. 2011). It also implies that we cannot obtain target energy input (or total energy input slack) from (3.7) directly. For this concern, Ali and Seiford (1993) propose a second stage LP problem to maximize the sum of non-radial slacks. Instead of the two-stage DEA method, moreover, Coelli (1998) suggests a multi-stage DEA method to acquire non-radial slacks.

Assume s_e^- is the non-radial slack of energy input calculated by either two- or multi-stage DEA. $(1–\theta) \times e_o$ is equal to the radial slack of energy input. Then total energy slack would be the sum of radial and non-radial slacks and target energy input is equal to actual energy input minus total energy slack. Hence, Hu and Wang's TFEE based on CCR model is represented by

$$
\text{CCR-type TFEE} = \frac{\theta \times e_o - s_e^-}{e_o} = 1 - \frac{(1-\theta) \times e_o + s_e^-}{e_o}. \tag{3.8}
$$

In line with original TFEE measure, Wei et al. (2009) investigate provincial energy efficiency in China. In addition, Hu and Kao (2007) apply the CCR-type TFEE to calculate energy-saving targets for APEC economies. With regard to BCC model, there is an additional constraint $\sum_{n=1}^{N} \lambda_n = 1$ which is appended to (3.7), indicating that the production frontier is allowed VRS technology. Unfortunately, to our best knowledge, none research studies TFEE using BCC model.

[4] CCR and BCC models can be represented in multiplier form or envelopment form. Coelli et al. (2005) suggest that the envelopment form involves fewer constraints than the multiplier form. Therefore, envelopment form is generally preferred to solve the linear programming problem and used in our study instead of multiplier form.

3.3.1.2 All Inputs Can Be Proportionally Adjusted with Different Ratios

As has been noted above, the weak efficiency problem in CCR model causes a difficulty in obtaining target energy input. Although Ali and Seiford (1993) and Coelli (1998) provide solutions to the problem, both of them have corresponding disadvantages (Coelli et al. 2005). Therefore, a group of DEA models which allows all inputs for proportional contracting with different ratios is built (Färe et al. 1994b). Rather than using a Farrell efficiency measure in CCR model, Färe and Lovell (1978) propose a famous DEA model based on a Russell efficiency measure.[5] Specifically, the Russell measure is well-behaved with less restrictive assumptions than Farrell measure (for details, see Färe and Lovell 1978). Thus, Russell efficiency measure usually has a higher discriminating power than Farrell efficiency measure.

According to our assumption mentioned above, we have a dataset consisting of $(M+1)$ inputs (energy consumption and M non-energy inputs). Hence, input-oriented Russell efficiency measure for the o-th DMU can be computed by following LP problem:

$$
\begin{aligned}
&\min \frac{1}{M+1}\left(\sum_{m=1}^{M}\theta_m+\theta_e\right)\\
&\text{s.t.} \quad \sum_{n=1}^{N}\lambda_n e_n \le \theta_e e_o,\\
&\qquad \sum_{n=1}^{N}\lambda_n x_{mn} \le \theta_m x_{mo}, \qquad m=1,2,\ldots,M,\\
&\qquad \sum_{n=1}^{N}\lambda_n y_{rn} \ge y_{ro}, \qquad r=1,2,\ldots,R,\\
&\qquad \lambda_n \ge 0, \quad n=1,2,\ldots,N.
\end{aligned}
\tag{3.9}
$$

Since (3.9) considers non-equally proportional adjustment of each input, this model can avoid the weak efficiency problem and obtain the strong efficiency scores. In terms of the concept of TFEE, the target energy input is defined as $\theta_e \times e_o$ and total energy input slack is equal to $(1-\theta_e)\times e_o$. Therefore, a Russell-type TFEE can be represented as:

[5] Other well-known extended DEA models with similar characteristics are the Zieschang measure and the asymmetric Färe measure (see De Borger and Kerstens 1996).

$$\text{Russell} - \text{type TFEE} = \frac{\theta_e \times e_o}{e_o} = 1 - \frac{(1-\theta_e) \times e_o}{e_o} = \theta_e. \tag{3.10}$$

It is noteworthy that Färe and Lovell's Russell efficiency measure looks for the minimum arithmetic mean of proportional decreases in all inputs. One can easily extend this model to a more general setting. For example, instead of the minimum arithmetic mean of object function in (3.9), Zhu (1996) suggests some weighted object functions that reflect the preference structure of DMUs.

There are several studies using the concept of Russell-type TFEE to investigate energy efficiency. For instance, Zhou and Ang (2008) construct an energy efficiency performance index based on Russell measure to evaluate the energy efficiency performances of 21 OECD countries. In addition, Hernández-Sancho et al. (2011) apply Russell measure to calculate energy efficiency of wastewater treatment plants in Spain.

3.3.1.3 Only Energy Can Be Radially Adjusted

There is a critical issue about the assumption of adjustment recently, i.e., can all inputs be decreased through either radial or non-radial adjustment? Theoretically, above-mentioned DEA models assume that all inputs can be freely and instantly adjusted to their target levels. However, in the real world, not all inputs can adjust to their optimal levels due to adjustment costs, regulation and indivisibilities (Ouellette and Vierstraete 2004). This is also a major concern as studying in energy efficiency field. From policy makers and the government's perspective, for example, no one is willing to cut employment level and reduce capital stocks along with decreasing excess energy consumption.

To deal with this issue, Ouellette and Vierstraete (2004) propose a modified DEA model under a quasi-fixed inputs assumption. With respect to energy efficiency study, the quasi-fixed inputs DEA model would set that only energy input can be adjusted and other inputs are quasi-fixed in the short-run. Under this setting, the o-th DMU's efficiency score can be solved by following LP problem:

$$\begin{aligned}
\min\ & \theta \\
\text{s.t.}\ & \sum_{n=1}^{N} \lambda_n e_n \le \theta e_o, \\
& \sum_{n=1}^{N} \lambda_n x_{mn} \le x_{mo}, \qquad m = 1, 2, \ldots, M, \\
& \sum_{n=1}^{N} \lambda_n y_{rn} \ge y_{ro}, \qquad r = 1, 2, \ldots, R, \\
& \lambda_n \ge 0, \quad n = 1, 2, \ldots, N.
\end{aligned} \tag{3.11}$$

Actually, (3.11) is very similar to CCR model shown in (3.7) but non-energy inputs are not allowed for adjusting in quasi-fixed inputs DEA model. Because this model does not radially adjust all inputs, moreover, the weak efficiency problem can be avoided. Therefore, based on the concept of TFEE, the target energy input is defined as $\theta \times e_o$ and total energy input slack is equal to $(1-\theta) \times e_o$. The TFEE under a quasi-fixed inputs assumption is

$$\text{QFI-type TFEE} = \frac{\theta \times e_o}{e_o} = 1 - \frac{(1-\theta) \times e_o}{e_o} = \theta. \tag{3.12}$$

Practically, quasi-fixed inputs assumption is closer to the reality than CCR model because labor and capital inputs may not be freely adjusted in short-run. As a result, for energy efficiency study, this measure has become very popular recently. For instance, Shi et al. (2010) consider a DEA model of fixing non-energy inputs to evaluate Chinese regional industrial energy efficiency. Zhou and Ang (2008) also introduce an energy efficiency performance index under a quasi-fixed inputs assumption.

In sum, this subsection briefly introduces three kinds of proportional adjustment DEA models and the relationship between these models and total-factor energy efficiency index. It is worth noting that the CCR-type TFEE is the easiest for understanding and can be computed through some free software, such as DEAP. However, as discussed above, there are two illogical assumptions of the CCR-type TFEE as for investigating energy issues. One is that why we can assume all inputs are adjusted with an equal weight and the other is that why we can assume all inputs can be freely adjusted. To deal with these shortcomings, the Russell-type TFEE relaxes the former assumption and the QFI-type TFEE further relaxes both of two assumptions indeed. In addition, the QFI-type and Russell-type TFEEs have higher discriminating powers than the CCR-type TFEE. Therefore, it is suggested that, in the future works, the CCR-type TFEE should not be used in energy efficiency field. Moreover, we would consider that the QFI-type is better than the Russell-type TFEE because the assumption of QFI-type TFEE is closer to the real world.

3.3.2 Slack-Based Models Without Undesirable Outputs

The second category of energy-related DEA models is traditional slack-based models (SBM) without undesirable outputs. Besides the Russell efficiency measure introduced above, the SBM models are also known as one famous kind of non-radial DEA models. Particularly, the SBM models directly deal with slacks, such as input excess and output shortfall, rather than proportional adjustment ratios. Therefore, according to the definition of TFEE index, the SBM models can calculate total energy slack and then find out target energy input more efficiently.

The term—slack-based model—is formally proposed by Tone (2001). In the earlier literature, Charnes et al. (1985) firstly introduce an additive DEA model rather than traditional radial model (such as CCR and BCC models) to account for all sources of inefficiency. Pastor et al. (1999) propose an enhanced Russell measure which is defined as a combination of the input and output Russell measures. They also express the corresponding formulas in terms of total slacks. Sequentially, with similar idea, Tone (2001) defines SBM DEA and shows that the SBM can be deemed as a product of input and output inefficiencies.

The original SBM DEA proposed by Tone (2001) is a non-oriented or both-oriented efficiency measure which deals with total slacks both in inputs and outputs. Hence, as noted in his paper, SBM-efficient is full efficient and presents a higher discriminating power than CCR model. According to the definition, non-oriented SBM can be formulated as the following fractional program:

$$\rho_{NO}^{*} = \min \frac{1 - \dfrac{1}{M+1}\left(\displaystyle\sum_{m=1}^{M} \frac{s_m^-}{x_{mo}} + \frac{s_e^-}{e_o}\right)}{1 - \dfrac{1}{R}\displaystyle\sum_{r=1}^{R} \frac{s_r^+}{y_{ro}}}$$

$$\begin{aligned} \text{s.t.} \quad & \sum_{n=1}^{N} \lambda_n e_n = e_o - s_e^-, \\ & \sum_{n=1}^{N} \lambda_n x_{mn} = x_{mo} - s_m^-, \qquad m = 1, 2, \ldots, M, \\ & \sum_{n=1}^{N} \lambda_n y_{rn} = y_{ro} + s_r^+, \qquad r = 1, 2, \ldots, R, \\ & \lambda_n \geq 0,\ s_e^- \geq 0,\ s_m^- \geq 0,\ s_r^+ \geq 0,\ n = 1, 2, \ldots, N. \end{aligned} \tag{3.13}$$

where ρ_{NO}^{*} is non-oriented SBM efficiency score. s_e^- is total slack in energy input; s_m^- and s_r^+ indicate the excess in the m-th input and shortfall in the r-th output, respectively. With respect to computational aspect, non-oriented SBM can be transformed into a linear program using Charnes and Cooper's transformation (Pastor et al. 1999; Tone 2001). However, the corresponding LP is inconvenient for obtaining TFEE score since the optimal solution of LP should be transferred to the optimal solution of fractional program mentioned above.

Accordingly, we introduce an easier way to calculate TFEE in SBM DEA as follows. Tone (2011) introduces a weighted-SBM DEA model which assigns weight to each slack variable in the objective function. The model is more flexible for researchers to allot relative importance of all inputs and outputs. In case we are only interested in energy efficiency (or total energy slack), we can assign all weight to energy input and then construct an input-oriented weighted-SBM DEA. Therefore, this input-oriented weighted-SBM DEA can be directly solved by the following LP problem:

$$
\begin{aligned}
&\rho_I^* = \min 1 - \left(s_e^-/e_o\right) \\
&\text{s.t.} \quad \sum_{n=1}^{N} \lambda_n e_n = e_o - s_e^-, \\
&\qquad \sum_{n=1}^{N} \lambda_n x_{mn} = x_{mo} - s_m^-, \qquad m = 1, 2, \ldots, M, \\
&\qquad \sum_{n=1}^{N} \lambda_n y_{rn} = y_{ro} + s_r^+, \qquad r = 1, 2, \ldots, R, \\
&\qquad \lambda_n \geq 0,\ s_e^- \geq 0,\ s_m^- \geq 0,\ s_r^+ \geq 0,\ n = 1, 2, \ldots, N.
\end{aligned}
\tag{3.14}
$$

Based on the definition in (3.2), ρ_I^* denotes not only input-oriented SBM efficiency but also TFEE score indeed. Hence, we define SBM-type TFEE as

$$
\text{SBM-type TFEE} = \rho_I^* = 1 - \frac{s_e^-}{e_o}. \tag{3.15}
$$

As discussed earlier, SBM-type TFEE can quickly tell us the amount of total energy slack (s_e^-) and target energy input ($e_o - s_e^-$). We consider that SBM-type TFEE would be more appropriate for policy makers because it intuitively points out how much energy is wasted. With respect to related literature, Zhou et al. (2006) establish their environmental performance indicator based on SBM efficiency measures. Choi et al. (2012) apply a SBM DEA model to calculate potential reduction of energy use for 30 provinces in China.

3.3.3 DEA Models with Energy Use and Pollutant Emissions

As discussed in Sect. 3.2, in the real production process, undesirable outputs would be produced together with desirable outputs. For example, Färe et al. (1996) investigate United States electric utilities' efficiency and indicate that these utilities are also a major contributor to sulfur dioxide, carbon dioxide and nitrogen oxide emissions. Hence, how to evaluate energy use and pollutant emissions simultaneously in a DEA model has become an attractive issue. Accordingly, this subsection will introduces the third group of energy-related DEA models that incorporate the reduction of energy input and pollutant outputs simultaneously.

With respect to undesirable outputs such as pollutant emissions, there are two theoretical concerns about the feasible technology in traditional DEA models. First, conventional DEA models only allow increases in outputs and decreases in inputs, meaning that decreases in undesirable outputs are restricted. To deal with this restriction, hence, some papers treat the undesirable outputs as inputs or use data translation to reverse the undesirable outputs in DEA models (Seiford and Zhu 2002).

Second, traditional DEA models assume that inputs and outputs are strongly (freely) disposable. However, the reduction of undesirable outputs should be costly in the real production process. Therefore, Färe et al. (1989) assume a weak disposability of undesirable outputs instead of strong disposability. Sequentially, Färe et al. (1996), Färe and Grosskopf (2004) and Färe et al. (2004) propose an environmental technology which imposes weak disposability and null-jointness properties on the undesirable outputs. Now, their works have been widely applied to evaluate energy or environmental efficiency as considering pollutant emissions (Zhou et al. 2008).

Regarding calculation aspect, we briefly introduce two commonly-used DEA models with undesirable outputs to compute TFEE. The first model proposed by Färe et al. (1996) is based on Shephard input distance function (DF). However, the original model assumes the proportional scaling of all inputs that would be similar to a Farrell efficiency measure. To approach the reality, we assume only energy input can be adjusted that is consistent with quasi-fixed inputs assumption. Hence, the o-th DMU's efficiency score can be solved by following LP problem:

$$
\begin{aligned}
& D_i(e_o, x_{mo}, y_{ro}, u_{ko})^{-1} = \min\ \rho \\
& \text{s.t.}\ \sum_{n=1}^{N} \lambda_n e_n \le \rho e_o, \\
& \sum_{n=1}^{N} \lambda_n x_{mn} \le x_{mo}, \qquad m = 1, 2, \ldots, M, \\
& \sum_{n=1}^{N} \lambda_n y_{rn} \ge y_{ro}, \qquad r = 1, 2, \ldots, R, \\
& \sum_{n=1}^{N} \lambda_n u_{kn} = u_{ko}, \qquad k = 1, 2, \ldots, K, \\
& \lambda_n \ge 0, \quad n = 1, 2, \ldots, N.
\end{aligned}
\tag{3.16}
$$

It is noted that (3.16) is quite similar to (3.11) introduced above. The only difference between these two models is that (3.16) copes with undesirable outputs imposed weak disposability and null-jointness properties. A ratio of the optimal solution of (3.16) to that of (3.11) is defined as an environmental performance indicator (Färe et al. 1996). Moreover, according to the definition of TFEE, therefore, the DF-type TFEE can be shown as:

$$
\text{DF-type TFEE} = \frac{\rho \times e_o}{e_o} = 1 - \frac{(1 - \rho) \times e_o}{e_o} = \rho. \tag{3.17}
$$

Aside from Shephard's distance function approach, Färe and Grosskopf (2004) use a more general approach—directional distance function proposed by Chung et al. (1997) to solve the LP problem. In fact, directional distance function (DDF) allows us to expand good outputs and contract undesirable outputs simultaneously. If we focus on an output-oriented concern, such as total-factor undesirable output efficiency, directional distance function must be more appropriate than Shephard's distance function. Turns to energy efficiency issue, with respect to input-oriented directional distance function approach, the o-th DMU's efficiency score can be solved by following optimization problem:

$$\begin{aligned}
&\vec{D}_i(e_o, x_{mo}, y_{ro}, u_{ko}; g_e) = \max\ \beta \\
&\text{s.t.}\ \ \sum_{n=1}^{N} \lambda_n e_n \le e_o - \beta g_e, \\
&\qquad \sum_{n=1}^{N} \lambda_n x_{mn} \le x_{mo}, \qquad m = 1, 2, \ldots, M, \\
&\qquad \sum_{n=1}^{N} \lambda_n y_{rn} \ge y_{ro}, \qquad r = 1, 2, \ldots, R, \\
&\qquad \sum_{n=1}^{N} \lambda_n u_{kn} = u_{ko}, \qquad k = 1, 2, \ldots, K, \\
&\qquad \lambda_n \ge 0, \ \ n = 1, 2, \ldots, N.
\end{aligned} \tag{3.18}$$

where g_e denote the directional vector for energy input. On one hand, if we take $g_e = e_o$, the optimal solution, β, can be interpreted as the proportional contraction in energy to stand onto the efficient frontier. On the other hand, if we take $g_e = 1$, β would be denoted the number of decreases in energy input, i.e., total energy slack. Therefore, for example, we set $g_e = e_o$ and then the DDF-type TFEE can be represented as the following equation:

$$DDF\text{-type TFEE} = \frac{e_o - \beta \times e_o}{e_o} = 1 - \beta. \tag{3.19}$$

It is obvious that the DF-type TFEE will be equal to the DDF-type TFEE under a quasi-fixed inputs assumption even though their objective functions are not the same. However, the setting of directional distance function is more general and distance function is a special case (Färe and Grosskopf 2004). In addition, one can apply slack-based DEA models to solve the LP problem with undesirable outputs and obtain its TFEE, such as the work of Zhou et al. (2006). But, Färe and Grosskopf (2010) introduce a generalized directional distance function approach and show that SBM DEA is a special case of the generalized DDF model. Because we adopt a quasi-fixed inputs assumption, the generalized DDF model should be in line with traditional DDF model. Then, the DDF-type TFEE will be consistent with the SBM-type one modelling undesirable outputs under a quasi-fixed inputs assumption.

3.4 An Illustration of Regional Energy Efficiency in China

Section 3.3 has introduced a variety of DEA models to calculate different types of TFEE indices. Accordingly, there are four types of TFEE indices without undesirable output, including CCR-type, Russell-type, QFI-type and SBM-type TFEE indices. In addition, we also present two types of TFEE indices with undesirable output, i.e., DF-type and DDF-type TFEE indices, respectively. Therefore, this section further uses Chinese regional data in 2010 to demonstrate and discuss the results among kinds of TFEEs. Following Hu and Wang (2006), we use four input variables consisting energy consumption, labor employment, capital stock and total sown area of farm crops. Two output variables are collected which are provincial GDP (desirable output) and sulfur dioxide emission (undesirable output) in this illustration. Table 3.1 lists data for our application including 29 DMUs using four inputs to produce two outputs. It is noted that 29 provinces and municipalities in China in 2010 are considered since Chongqing is regarded as a part of Sichuan and Tibet's energy consumption is unavailable.

Table 3.2 shows the computation results for six TFEE indices. Actually, we can discuss the results from three aspects. First, among columns two to five (DEA models without undesirable outputs), the number of efficient DMUs is the lowest by using Russell-type TFEE index. It is not surprising that Russell-type TFEE has the highest discriminating power since it allows all inputs can be adjusted with non-equally proportions. Second, as we compute the range of each TFEE index (one minus the minimum TFEE), QFI- and SBM-type TFEEs present the largest range among DMUs' TFEE. It means that QFI- and SBM-type TFEEs still have higher discriminating power although the number of efficient DMUs using these indices is not the smallest. Notice that under a quasi-fixed inputs assumption, TFEEs obtained from QFI and SBM DEA models will be exactly the same. As a result, we can conclude that CCR model may overestimate DMUs' TFEE performance.

Third, after we take into account SO_2 emission (the last two columns of Table 3.2), the number of efficient DMUs increases and the range of TFEE decreases. It is suggested that the discriminating power of DF- and DDF-type TFEEs are lower than other models without SO_2 emission. The gap between efficient and inefficient Chinese provinces narrows when SO_2 emission is denoted as an undesirable output. One possible explanation is that we may overestimate (underestimate) the TFEE of efficient (inefficient) provinces if we neglect pollutant outputs in DEA models. This finding is consistent with the work of Li and Hu (2012). Furthermore, this result implies that total-factor undesirable output efficiency would be a critical challenge in China.

Table 3.1 Data of regions in China in 2010

ID	GDP (100 million RMB)	SO_2 (ton)	Energy (10,000 tce)	Labor (10,000 person)	Capital (100 million RMB)	Farm area (1000 ha)
Beijing	14,114	115,050	6954	970	2254	317
Tianjin	9224	235,150	6818	329	1072	459
Hebei	20,394	1,233,780	27,531	814	2406	8718
Shanxi	9201	1,249,201	16,808	565	756	3764
Inner Mongolia	11,672	1,394,100	16,820	465	2156	7003
Liaoning	18,457	1,022,207	20,947	1030	955	4074
Jilin	8668	356,310	8297	515	1056	5221
Heilongjiang	10,369	490,164	11,234	754	692	12,156
Shanghai	17,166	358,100	11,201	736	7764	401
Jiangsu	41,425	1,050,488	25,774	2061	4492	7620
Zhejiang	27,722	678,342	16,865	1642	1996	2485
Anhui	12,359	532,076	9707	770	265	9053
Fujian	14,737	409,051	9809	786	776	2271
Jiangxi	9451	557,072	6355	545	1340	5458
Shandong	39,170	1,537,818	34,808	1593	4334	10,818
Henan	23,092	1,338,701	21,438	1127	2065	14,249
Hubei	15,968	632,582	15,138	962	898	7998
Hunan	16,038	801,311	14,880	878	660	8216
Guangdong	46,013	1,050,508	26,908	2352	2734	4525
Guangxi	9570	903,826	7919	558	649	5897
Hainan	2065	28,810	1359	161	97	834
Sichuan	25,111	1,850,356	25,748	1562	747	12,838
Guizhou	4602	1,148,830	8175	324	264	4889
Yunnan	7224	500,702	8674	647	78	6437
Shaanxi	10,123	778,649	8882	482	1332	4186
Gansu	4121	551,785	5923	318	1190	3995
Qinghai	1350	143,431	2568	96	112	547
Ningxia	1690	310,752	3681	108	169	1248
Xinjiang	5437	588,487	8290	387	507	4759

Note: (1) All monetary values are 2010 prices; (2) In fact, the analyzed data is rounded off to the second digit after the decimal point

We further compute the Spearman's rank correlation among six models and present the result in Table 3.3. Four TFEE indices without SO_2 emission highly and positively correlate to each other, indicating that the rank of provincial TFEE among models is accordant even though these models suppose different assumptions. In addition, the correlation between TFEE indices with and without SO_2 emission is weaker (around 0.5), showing that the rank is shuffled as we consider pollutant outputs in DEA models.

Table 3.2 Total-factor energy efficiency scores of Chinese regions in 2010 for six measures

ID	Area	CCR	Russell	QFI	SBM	DF	DDF
Beijing	E	1.000	1.000	1.000	1.000	1.000	1.000
Tianjin	E	1.000	1.000	1.000	1.000	1.000	1.000
Hebei	E	0.559	0.433	0.516	0.516	0.690	0.690
Shanxi	C	0.454	0.320	0.309	0.309	1.000	1.000
Inner Mongolia	W	0.513	0.513	0.482	0.482	1.000	1.000
Liaoning	E	0.609	0.515	0.555	0.555	1.000	1.000
Jilin	C	0.726	0.611	0.566	0.566	0.648	0.648
Heilongjiang	C	0.732	0.540	0.534	0.534	0.614	0.614
Shanghai	E	1.000	0.755	1.000	1.000	1.000	1.000
Jiangsu	E	0.977	0.940	0.959	0.959	0.977	0.977
Zhejiang	E	0.989	0.961	0.942	0.942	1.000	1.000
Anhui	C	1.000	1.000	1.000	1.000	1.000	1.000
Fujian	E	1.000	0.879	1.000	1.000	1.000	1.000
Jiangxi	C	0.877	0.870	0.819	0.819	1.000	1.000
Shandong	E	0.864	0.658	0.775	0.775	0.846	0.846
Henan	C	0.890	0.630	0.654	0.654	0.816	0.816
Hubei	C	0.878	0.617	0.634	0.634	0.695	0.695
Hunan	C	1.000	0.630	1.000	1.000	1.000	1.000
Guangdong	E	1.000	1.000	1.000	1.000	1.000	1.000
Guangxi	W	0.870	0.707	0.697	0.697	1.000	1.000
Hainan	E	0.990	0.888	0.986	0.986	1.000	1.000
Sichuan	W	1.000	0.649	1.000	1.000	1.000	1.000
Guizhou	W	0.518	0.329	0.336	0.336	1.000	1.000
Yunnan	W	1.000	1.000	1.000	1.000	1.000	1.000
Shaanxi	W	0.809	0.667	0.703	0.703	1.000	1.000
Gansu	W	0.495	0.407	0.343	0.343	0.744	0.744
Qinghai	W	0.406	0.308	0.296	0.296	0.614	0.614
Ningxia	W	0.389	0.268	0.251	0.251	0.724	0.724
Xinjiang	W	0.571	0.384	0.363	0.363	0.598	0.598
# of Efficient DMU		9	5	9	9	18	18

Note: E, W and C indicate east, west and central areas in China, respectively

Table 3.3 Spearman rank correlation matrix for six efficiency measures

	CCR	Russell	QFI	SBM	DF	DDF
CCR	1.000					
Russell	0.877	1.000				
QFI	0.981	0.901	1.000			
SBM	0.981	0.901	1.000	1.000		
DF	0.533	0.558	0.587	0.587	1.000	
DDF	0.533	0.558	0.587	0.587	1.000	1.000

3.5 Further Extended Topics

3.5.1 Selection of Input and Output Variables

The first related topic is how to select relevant inputs and outputs. In the review paper, Zhou et al. (2008) provide some suggestions about screening procedures and the number of input and output variables. However, there has been a growing body of researches investigating energy efficiency issue over the past decades. Due to their diverse research objects, there is no unique criterion to select inputs and outputs. Therefore, we collect around 30 application papers published in famous energy-related journals after 2008. It is noted that these articles investigate either energy efficiency or undesirable output efficiency performance. We further classify these articles based on the following attributes: type of DMUs, output and input variables, and DEA models. As a result, the features of collected papers are listed in Table 3.4.

According to Table 3.4, we can briefly summarize two conclusions about the selection of input and output variables. First, output variables are clearly different on the basis of type of DMUs. More specifically, in case the studies apply country or regional (i.e., provinces, states and prefectures) level data, economic outputs should be better than physical outputs. However, if the DMUs are plants, farms or factories, physical outputs would be more appropriate, such as electricity production and crop yield. Second, whatever types of DMUs are analyzed, all of labor force, capital stock and energy consumption should be contained in the inputs list if possible. Because both of labor and capital are traditional input factors in the theory of production, they are essentials to evaluate total-factor energy efficiency index indeed. It is worth noting that a few researches only use energy consumption as input variable due to their limited number of DMUs (e.g., Sueyoshi and Goto 2013; Zhou et al. 2008).

3.5.2 Total-Factor Energy Productivity Growth

The second extended topic discusses about the dynamic analysis of TFEE index, i.e., total-factor energy productivity growth. As introduced above, TFEE compares DUMs' relative energy efficiency at particular one time point. Chang and Hu (2010) argue that TFEE cannot completely depict the dynamic pattern of energy usage due to the neglect of technical change. Hence, we briefly introduce three commonly-used productivity indices to compute total-factor energy productivity change.

The first well-known index is the Malmquist productivity index. Färe et al. (1994a) apply DEA methodology with output distance function to estimate the Malmquist productivity index. Färe et al. (1994c) further decompose Malmquist productivity index into two components, i.e., technical change and efficiency change. In terms of energy productivity growth, however, an input-oriented

Table 3.4 Literature of DEA models with their specific features

Authors/Year	Type of DMUs	Output	Inputs	Method
Banaeian and Zangeneh (2011)	Iranian corn producers	Corn yield	Labor, machinery, diesel fuel, chemical fertilizers, farmyard manure, chemicals, water, seeds	CCR
Bian and Yang (2010)	Chinese provinces	GDP, COD, Nitrogen, SO_2	Labor, capital, energy, water	SBM, DDF
Blomberg et al. (2012)	Swedish pulp and paper mills	Output of pulp and paper	Labor, oil, electricity	CCR
Chang and Hu (2010)	Chinese provinces	GDP	Capital, labor, energy, farm area	DDF
Chang et al. (2013)	Chinese transportation sector in provincial level	CO_2 emissions, value-added	Labor, fixed capital investment, energy consumption	SBM
Choi et al. (2012)	Chinese provinces	GDP, CO_2 emission	Capital, labor, energy	SBM
Fallahi et al. (2011)	Iranian power electric generation companies	Net electricity produced	Employees, fuel, electricity used, capacity, operational time	CCR
Guo et al. (2011)	Chinese provinces	GDP, CO_2 emission	Capital, labor, energy used	DF, SBM
Hernández-Sancho et al. (2011)	Spanish wastewater treatment plants	Suspended solids, COD	Energy, staff, reagents, waste management, maintenance, other	Russell
Honma and Hu (2008)	Japanese prefectures	Total income	11 energy inputs, labor, private and public capital	CCR
Honma and Hu (2009)	Japanese prefectures	Total income	11 energy inputs, labor, private and public capital	DF
Iribarren, et al. (2013)	Spanish wind farms	Electricity	Concrete, steel, working hours, epoxy resin, oil	SBM
Khoshroo et al. (2013)	Iranian grape farmers	Grape yield	Labor, machinery, chemicals, farmyard manure, diesel, electricity, water	CCR, BCC
Li and Hu (2012)	Chinese provinces	GDP, CO_2, SO_2	Labor, capital, energy	SBM
Liu et al. (2010)	Thermal power plants in Taiwan	Net electricity produced	Electricity used, heating value of total fuels, installed capacity	CCR
Mandal and Madheswaran (2010)	Indian cement industry in state level	Value of ex-factory products, CO_2	Labor, capital, energy, materials	DDF
Mousavi-Avval et al. (2011)	Iranian apple farmers	Apple yield	Labor, machinery, diesel fuel, chemicals, FYM, chemical fertilizer, water, electricity	CCR

(continued)

Table 3.4 (continued)

Authors/Year	Type of DMUs	Output	Inputs	Method
Mukherjee (2008a)	Indian manufacturing sector in state level	Gross value of manufacturing production	Labor, capital, energy, materials	CCR, QFI
Mukherjee (2008b)	U.S. manufacturing sectors	Gross output	Labor, capital, energy, materials, services	QFI
Shi et al. (2010)	Chinese industry in provincial level	Industrial value added, industrial waste gas	Fixed assets, industrial energy consumption, labor,	QFI
Song et al. (2013)	BRICS countries	GDP	Labor, capital formation rate, energy consumption	CCR
Sueyoshi and Goto (2011)	Japanese Fossil fuel power generations	Power generation, CO_2 emission	Coal, oil, LNG, employees, generation capacity	QFI, SBM
Sueyoshi and Goto (2012)	U.S. coal-fired power plants	Annual net generation, SO_2, NO_X, CO_2	Employees, total cost of plant, total non-fuel O&M, fuel consumption	SBM
Sueyoshi and Goto (2013)	Ten industrial countries	Electricity, CO_2	Combustible, nuclear, Hydro + renewables	SBM
Wu et al. (2012)	Chinese industry in provincial level	Industrial value added, CO_2	Capital, labor, energy	DF
Zhang et al. (2011)	23 developing countries	GDP	Labor, energy, capital stock	CCR
Zhang et al. (2013)	Korean fossil fuel electricity generation	Gross electricity generation, CO_2 emissions	Capital, fossil fuel, labor	DDF
Zhou and Ang (2008)	21 OECD countries	GDP, CO_2	Capital stock, labor force, coal, oil, gas, other energy	Russell, QFI, DF, SBM
Zhou et al. (2008)	Eight world regions	GDP, CO_2 emissions	Energy consumption	DF
Zhou et al. (2012)	Electricity generation of 129 countries	Electricity generated, CO_2 emissions	fossil fuel consumption	DDF
Zhou et al. (2013)	Chinese power industry	Electricity production, SO_2, NO_X, CO_2	Energy consumption, labor, fixed assets investment	SBM

Malmquist productivity index should be more appropriate than an output-oriented one. Honma and Hu (2009) propose a total-factor energy productivity index through using an input-oriented Malmquist productivity index. With respect to computational aspect, their proposed index is solved by (3.7) and (3.8) introduced above that assumes all inputs can be radial adjusted. Of course, under a quasi-fixed inputs assumption, one can use (3.11) to solve corresponding input distance functions. Besides, Zhou et al. (2010) focus on the productivity growth of undesirable output and introduce a total-factor carbon performance measure in the same manner.

In fact, there is a major problem in Malmquist productivity index if undesirable outputs are presented in DEA models. That is, Shephard's distance function cannot satisfy the environmental technology that seeks to increase desirable outputs and decrease undesirable outputs simultaneously. Therefore, Chambers et al. (1996) construct a Luenberger productivity index based on directional distance functions which can be solved by DEA methodology.[6] Regards to energy issue, Chang and Hu (2010) introduce a total-factor energy productivity index using the Luenberger productivity index to analyze the energy productivity growth in China. Although their index assumes all inputs can be radial adjusted, one can solve the corresponding input directional distance functions by aforementioned (3.18) under a quasi-fixed inputs assumption.

The third index, so-called Malmquist-Luenberger productivity index proposed by Chung et al. (1997), combines the features of the Malmquist and Luenberger productivity indices. Specifically, the Malmquist-Luenberger index is based on directional distance functions and has both of multiplicative and additive structures; see Chung et al. (1997) for details. Because of the well behavioral assumption, the Malmquist-Luenberger index has been adopted in evaluating environmental productivity with undesirable outputs. Recently, both He et al. (2013) and Wang et al. (2013) apply a total-factor Malmquist-Luenberger energy productivity to investigate the energy productivity issue in China. In addition, since the Malmquist-Luenberger index is constructed by directional distance functions, one can solve those functions by using foregoing (3.18) indeed.

Finally, we also present an application of energy productivity change using Chinese provincial data from 2010 to 2011. It is noted that input and output variables are identical to Table 3.1 and all monetary values are 2010 prices.[7] The results obtained for the Malmquist, Luenberger and Malmquist-Luenberger indices are shown in Table 3.5.

Column two of Table 3.5 reports total-factor energy productivity changes using the Malmquist index without considering SO_2 emission. Accordingly, 23 of 29 provinces present positive energy productivity growth during the period 2010–2011. Sichuan has the highest total-factor energy productivity growth rate (80.1 %), while Qinghai is the worst one (−25.2 %). However, as shown in the last two columns, the results obtained for Luenberger and Malmquist-Luenberger indices are quite different from Malmquist index. If we consider SO_2 emission in DEA models, only 11 provinces represent increases in total-factor energy productivity. Hunan is the worst performer, while the growth rate of Sichuan becomes a

[6] Due to the flexibility of directional distance function, Boussemart et al. (2003) suggest that the Luenberger productivity index is more appropriate than the Malmquist productivity index.

[7] The Chinese provincial data in 2011 is available on request to author.

Table 3.5 Total-factor energy productivity growth for three indices of regions in China (2010–2011)

ID	Area	Malmquist	Luenberger	Malmquist-Luenberger
Beijing	E	0.038	−0.071	−0.075
Tianjin	E	0.021	−0.024	−0.025
Hebei	E	0.168	0.106	0.126
Shanxi	C	0.230	−0.013	−0.015
Inner Mongolia	W	0.185	−0.029	−0.030
Liaoning	E	0.796	0.055	0.062
Jilin	C	0.317	0.038	0.059
Heilongjiang	C	0.255	0.028	0.045
Shanghai	E	0.076	−0.091	−0.096
Jiangsu	E	0.317	0.027	0.028
Zhejiang	E	0.278	−0.039	−0.040
Anhui	C	0.243	−0.010	−0.010
Fujian	E	0.168	−0.043	−0.044
Jiangxi	C	−0.099	−0.023	−0.021
Shandong	E	0.188	0.071	0.082
Henan	C	0.094	−0.009	−0.010
Hubei	C	0.438	0.030	0.043
Hunan	C	−0.187	−0.174	−0.174
Guangdong	E	0.001	−0.120	−0.118
Guangxi	W	0.175	−0.144	−0.142
Hainan	E	0.376	−0.079	−0.082
Sichuan	W	0.801	−0.067	−0.070
Guizhou	W	0.457	−0.142	−0.124
Yunnan	W	0.020	0.071	0.079
Shaanxi	W	−0.238	0.000	0.000
Gansu	W	0.108	0.023	0.030
Qinghai	W	−0.252	−0.017	−0.027
Ningxia	W	−0.058	0.147	0.188
Xinjiang	W	−0.152	0.030	0.047

Note: E, W and C indicate east, west and central areas in China, respectively

negative number. In addition, only 12 provinces have the same sign of energy productivity changes among three kinds of measures. These results echo the conclusion of Table 3.3 that we may overestimate the energy productivity changes in China if we ignore pollutant outputs in DEA models.

3.6 Concluding Remarks

Rather than traditional partial factor energy efficiency index, this paper introduces the concept of total-factor energy efficiency (TFEE) index and how to compute it through DEA methodology. This paper presents six well-known DEA models with different assumptions and discusses the link between TFEE and these models. Specifically, four of them are standard DEA models which do not consider undesirable outputs, while other two models assume environmental technology with undesirable outputs in order to meet the real energy consumption process. Moreover, an application of China's energy efficiency using each DEA model is illustrated in this paper. It is found that the TFEE index is sensitive to whether undesirable outputs are presented in DEA models. We suggest that this finding is reasonable since undesirable outputs are inevitably produced along with energy consumption.

This paper further discusses two extended topics in energy efficiency research. First, we review around 30 pieces of literature published recently and summarize how to select relevant input and output variables. Second, with respect to a dynamic analysis, we demonstrate three commonly-used productivity indices with or without undesirable outputs to calculate total-factor energy productivity growth between two periods. Therefore, it is believed that this paper is useful for researchers and policy makers who have interest in energy efficiency issues.

References

Ali AI, Seiford LM (1993) The mathematical programming approach to efficiency analysis. In: Fried HO, Lovell CAK, Schmidt SS (eds) The measurement of productive efficiency: techniques and applications. Oxford University Press, New York, pp 120–159

Banaeian N, Zangeneh M (2011) Study on energy efficiency in corn production of Iran. Energy 36:5394–5402

Banker RD, Charnes A, Cooper WW (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis. Manage Sci 30:1078–1092

Bian Y, Yang F (2010) Resource and environment efficiency analysis of provinces in China: a DEA approach based on Shannon's entropy. Energ Policy 38:1909–1917

Blomberg J, Henriksson E, Lundmark R (2012) Energy efficiency and policy in Swedish pulp and paper mills: a data envelopment analysis approach. Energ Policy 42:569–579

Boussemart JP, Briec W, Kerstens K, Poutineau JC (2003) Luenberger and Malmquist productivity indices: theoretical comparisons and empirical illustration. Bull Econ Res 55:391–405

Chang TP, Hu JL (2010) Total-factor energy productivity growth, technical progress, and efficiency change: an empirical study of China. Appl Energy 87:3262–3270

Chang YT, Zhang N, Danao D, Zhang N (2013) Environmental efficiency analysis of transportation system in China: a non-radial DEA approach. Energ Policy 58:277–283

Chambers RG, Färe R, Grosskopf S (1996) Productivity growth in APEC countries. Pac Econ Rev 1:181–190

Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision making units. Eur J Oper Res 2:429–444

Charnes A, Cooper WW, Golany B, Seiford LM, Stutz J (1985) Foundations of data envelopment analysis for Pareto-Koopmans efficient empirical production functions. J Econometrics 30:91–107

Choi Y, Zhang N, Zhou P (2012) Efficiency and abatement costs of energy-related CO_2 emissions in China: a slacks-based efficiency measure. Appl Energy 98:198–208

Chung YH, Färe R, Grosskopf S (1997) Productivity and undesirable outputs: a directional distance function approach. J Environ Manage 51:229–240

Coelli TJ (1998) A multi-stage methodology for the solution of oriented DEA models. Oper Res Lett 23:143–149
Coelli TJ, Rao DSP, O'Donnell CJ, Battese GE (2005) An introduction to efficiency and productivity analysis, 2nd edn. Springer, New York
Cooper WW, Seiford LM, Zhu J (2011) Data envelopment analysis: history, models, and interpretations. In: Cooper WW, Seiford LM, Zhu J (eds) Handbook on data envelopment analysis, 2nd edn. Springer, New York, pp 1–39
De Borger B, Kerstens K (1996) Radial and nonradial measures of technical efficiency: an empirical illustration for Belgian local government. J Product Anal 7:41–62
Fallahi A, Ebrahimi R, Ghaderi SF (2011) Measuring efficiency and productivity change in power electric generation management companies by using data envelopment analysis: a case study. Energy 36:6398–6405
Färe R, Grosskopf S (2004) Modeling undesirable factors in efficiency evaluation: comment. Eur J Oper Res 157:242–245
Färe R, Grosskopf S (2010) Directional distance functions and slacks-based measures of efficiency. Eur J Oper Res 200:320–322
Färe R, Grosskopf S, Hernandez-Sancho F (2004) Environmental performance: an index number approach. Resour Energy Econ 26:343–352
Färe R, Grosskopf S, Lindgren B, Roos P (1994a) Productivity developments in Swedish hospitals: a malmquist output index approach. In: Charnes A, Cooper WW, Lewin A, Seiford LM (eds) Data envelopment analysis: theory, methodology and applications. Kluwer Academic, Boston, pp 253–272
Färe R, Grosskopf S, Lovell CAK (1994b) Production frontier. Cambridge University Press, Cambridge
Färe R, Grosskopf S, Lovell CAK, Pasurka C (1989) Multilateral productivity comparisons when some outputs are undesirable: a nonparametric approach. Rev Econ Statist 71:90–98
Färe R, Grosskopf S, Norris M, Zhang Z (1994c) Productivity growth, technical progress and efficiency change in industrialized countries. Am Econ Rev 84:66–83
Färe R, Grosskopf S, Tyteca D (1996) An activity analysis model of the environmental performance of firms—application to fossil-fuel-fired electric utilities. Ecol Econ 18:161–175
Färe R, Lovell CAK (1978) Measuring the technical efficiency of production. J Econ Theory 19:150–162
Farrell MJ (1957) The measurement of productive efficiency. J R Statist Soc Ser A 120:253–290
Guo XD, Zhu L, Fan Y, Xie BC (2011) Evaluation of potential reductions in carbon emissions in Chinese provinces based on environmental DEA. Energ Policy 39:2352–2360
He F, Zhang Q, Lei J, Fu W, Xu X (2013) Energy efficiency and productivity change of China's iron and steel industry: accounting for undesirable outputs. Energ Policy 54:204–213
Hernández-Sancho F, Molinos-Senante M, Sala-Garrido R (2011) Energy efficiency in Spanish wastewater treatment plants: a non-radial DEA approach. Sci Total Environ 409:2693–2699
Honma S, Hu JL (2008) Total-factor energy efficiency of regions in Japan. Energ Policy 36:821–833
Honma S, Hu JL (2009) Total-factor energy productivity growth of regions in Japan. Energ Policy 37:3941–3950
Hu JL, Kao CH (2007) Efficient energy-saving targets for APEC economies. Energ Policy 35:373–382
Hu JL, Wang SC (2006) Total-factor energy efficiency of regions in China. Energ Policy 34:3206–3217
Hu JL, Wang SC, Yeh FY (2006) Total-factor water efficiency of regions in China. Res Policy 31:217–230
Iribarren D, Martín-Gamboa M, Dufour J (2013) Environmental benchmarking of wind farms according to their operational performance. Energy 61:589–597
Khoshroo A, Mulwa R, Emrouznejad A, Arabi B (2013) A non-parametric data envelopment analysis approach for improving energy efficiency of grape production. Energy 63:189–194

Li LB, Hu JL (2012) Ecological total-factor energy efficiency of regions in China. Energ Policy 46:216–224

Liu CH, Lin SJ, Lewis C (2010) Evaluation of thermal power plant operational performance in Taiwan by data envelopment analysis. Energ Policy 38:1049–1058

Mandal SK, Madheswaran S (2010) Environmental efficiency of the Indian cement industry: an interstate analysis. Energ Policy 38:1108–1118

Mousavi-Avval SH, Rafiee S, Mohammadi A (2011) Optimization of energy consumption and input costs for apple production in Iran using data envelopment analysis. Energy 36:909–916

Mukherjee K (2008a) Energy use efficiency in the Indian manufacturing sector: an interstate analysis. Energ Policy 36:662–672

Mukherjee K (2008b) Energy use efficiency in U.S. manufacturing: a nonparametric analysis. Energy Econ 30:76–96

Ouellette P, Vierstraete V (2004) Technological change and efficiency in the presence of quasi-fixed inputs: a DEA application to the hospital sector. Eur J Oper Res 154:755–763

Pastor JT, Ruiz JL, Sirvent I (1999) An enhanced DEA Russell graph efficiency measure. Eur J Oper Res 115:596–607

Seiford LM, Zhu J (2002) Modeling undesirable factors in efficiency evaluation. Eur J Oper Res 142:16–20

Shi G-M, Bi J, Wang J-N (2010) Chinese regional industrial energy efficiency evaluation based on a DEA model of fixing non-energy inputs. Energ Policy 38:6172–6179

Song ML, Zhang LL, Liu W, Fisher R (2013) Bootstrap-DEA analysis of BRICS' energy efficiency based on small sample data. Appl Energy 112:1049–1055

Sueyoshi T, Goto M (2011) DEA approach for unified efficiency measurement: assessment of Japanese fossil fuel power generation. Energy Econ 33:292–303

Sueyoshi T, Goto M (2012) Environmental assessment by DEA radial measurement: U.S. coal-fired power plants in ISO (Independent System Operator) and RTO (Regional Transmission Organization). Energy Econ 34:663–676

Sueyoshi T, Goto M (2013) DEA environmental assessment in a time horizon: malmquist index on fuel mix, electricity and CO_2 of industrial nations. Energy Econ 40:370–382

Tone K (2001) A slack-based measure of efficiency in data envelopment analysis. Eur J Oper Res 130:498–509

Tone K (2011) Slacks-based measure of efficiency. In: Cooper WW, Seiford LM, Zhu J (eds) Handbook on data envelopment analysis, 2nd edn. Springer, New York, pp 195–209

Wang H, Zhou P, Zhou DQ (2013) Scenario-based energy efficiency and productivity in China: a non-radial directional distance function analysis. Energy Econ 40:795–803

Wei C, Ni J, Shen M (2009) Empirical analysis of provincial energy efficiency in China. China World Econ 17:88–103

Wu F, Fan LW, Zhou P, Zhou DQ (2012) Industrial energy efficiency with CO_2 emissions in China: a nonparametric analysis. Energ Policy 49:164–172

Zhang N, Zhou P, Choi Y (2013) Energy efficiency, CO2 emission performance and technology gaps in fossil fuel electricity generation in Korea: a meta-frontier non-radial directional distance function analysis. Energ Policy 56:653–662

Zhang XP, Cheng XM, Yuan JH, Gao XJ (2011) Total-factor energy efficiency in developing countries. Energ Policy 39:644–650

Zhou P, Ang BW (2008) Linear programming models for measuring economy-wide energy efficiency performance. Energ Policy 36:2911–2916

Zhou P, Ang BW, Han JY (2010) Total factor carbon emission performance: a Malmquist index analysis. Energy Econ 32:194–201

Zhou P, Ang BW, Poh KL (2006) Slacks-based efficiency measures for modeling environmental performance. Ecol Econ 60:111–118

Zhou P, Ang BW, Poh KL (2008) A survey of data envelopment analysis in energy and environmental studies. Eur J Oper Res 189:1–18

Zhou P, Ang BW, Wang H (2012) Energy and CO2 emission performance in electricity generation: a non-radial directional distance function approach. Eur J Oper Res 221:625–635
Zhou Y, Xing X, Fang K, Liang D, Xu C (2013) Environmental efficiency analysis of power industry in China based on an entropy SBM model. Energ Policy 57:68–75
Zhu J (1996) Data envelopment analysis with preference structure. J Oper Res Soc 47:136–150

Chapter 4
Social Cost Efficient Service Quality: Integrating Customer Valuation in Incentive Regulation—Evidence from the Case of Norway

Christian Growitsch, Tooraj Jamasb, Christine Müller, and Matthias Wissner

Abstract In order to overcome the perverse incentives of excessive maintenance reductions and insufficient network investments arising with incentive regulation of electricity distribution companies, regulators throughout Europe have started regulating service quality. In this paper, we explore the impact of incorporating customers' willingness-to-pay for service quality in benchmarking models on cost efficiency of distribution networks. Therefore, we examine the case of Norway, which features this approach to service quality regulation. We use the data envelopment analysis technique to analyse the effectiveness of such regulatory instruments. Moreover, we discuss the extent to which this indirect regulatory instrument motivates a socially desired service quality level. The results indicate that internalising external or social cost of service quality does not seem to have played an important role in improving cost efficiency in Norwegian distribution utilities.

Keywords Electricity distribution • Service quality • Willingness-to-pay • Data envelopment analysis

C. Growitsch (✉)
University of Hamburg, Hamburg, Germany
e-mail: christian.growitsch@uni-hamburg.de

T. Jamasb
Durham University Business School, Durham, UK
e-mail: tooraj.jamasb@durham.ac.uk

C. Müller • M. Wissner
Department Energy Markets and Energy Regulation, WIK, Rhoendorfer Str. 68, 53604, Bad Honnef, Germany

J. Zhu (ed.), *Data Envelopment Analysis*, International Series in Operations Research & Management Science 238, DOI 10.1007/978-1-4899-7684-0_4

4.1 Introduction

The transition from cost-plus to incentive regulation of natural monopoly energy networks entails numerous new challenges for regulators and network operators. In principle, the objective of incentive regulation is to encourage network operators to improve their cost efficiency towards a given target and to reward them for over-performance and penalise them for under-performance. The underlying parameter is a regulatory formula that caps the allowed prices (price cap regulation) or the allowed revenues (revenue cap regulation) of a network operator. This stimulus may, however, create perverse incentives with regard to the level of supplied service quality. The network operator may focus solely on efficiency targets to the detriment of maintaining an adequate level of quality. Therefore, service quality regulation is being introduced in a growing number of countries.

Quality in the electricity distribution and retail sector is a multi-faceted output that comprises technical and non-technical dimensions. The aspects that are usually regulated span three main areas: commercial quality, voltage quality, and continuity of supply and/or reliability (CEER 2008). The generic terms for these three dimensions are "service quality" and "service quality regulation", respectively. In this paper we focus on the aspect of continuity of supply in distribution networks, which is arguably the most important and widely targeted dimension of service quality by regulators (see e.g., CEER 2008). The aspect of commercial quality in the retail sector as well as the technical issue of voltage quality is not part of our study.

From a regulatory point of view, continuity of supply appears in two dimensions: the first dimension is its availability to energy to customers (or inversely the absence of interruptions). Basically, this dimension can be measured by different (groups of) indicators,[1] either the customer minutes lost (e.g., in form of the SAIDI[2]), the number of interruptions (e.g., in form of the SAIFI[3]) or the energy not supplied (ENS), which gives the total amount of energy that would have been supplied to a customer if there would not have been any interruption. The second regulatory dimension is the customers' preference for continuity of supply. One form to measure customer preferences is to reveal their willingness-to-pay (WTP) for a certain service quality level, or for its inverse, i.e., the interruptions cost (IC) customers incur due to poor quality (Fumagalli et al. 2007).

With regard to the latter dimension, incentive based penalty and reward schemes for continuity of supply performance prove to be a sophisticated regulatory instrument to excite the regulated company to deliver a desired service quality level to its

[1] For a detailed overview of the different indicators employed in European countries, please refer to CEER (2008).

[2] System Average Interruption Duration Index, which gives the amount of time per year that the supply to a customer is interrupted.

[3] System Average Interruption Frequency Index, which gives the average number of time per year that the supply to a customer is interrupted.

customers.[4] More specifically, these schemes adjust the companies' revenues according to their performance against a predefined service quality indicator, e.g., the ENS[5] or the SAIDI combined with the customer's WTP.[6] For the company, higher quality levels are associated with higher revenues and vice-versa.

Whilst the UK, the Scandinavian countries, the Netherlands, Italy and Portugal can be considered as pioneers in this field in Europe, more and more other European regulators follow this practise and enhance or introduce regulatory approaches[7] for incentive based service quality regulation. When introducing such a scheme, the objective usually is to neutralize potential quality deterioration due to incentive regulation. Mostly, a baseline is defined to move towards the desired level of service quality. As a matter of fact, the service quality level in Europe substantially differs among the different countries and regulators employ diverse approaches to measure it. Overall, there are rather decreasing trends for customer minutes lost and the trends indicate a rapprochement of the continuity level in Europe (CEER 2008).

From a national perspective, however, the main regulatory challenge is to define the accurate reference for the country specific, socially desired level of continuity of supply rather than targeting regulatory measures towards a maximum quality level. Therefore, the regulator needs to know the companies cost of providing service quality as well as the customers' preferences. Provided with that, the pivotal regulatory objective is to harmonize the utilities' profit incentive with economic efficiency and customers' preferences in terms of continuity of supply. In other words, the idea is to internalize the costs of (poor) quality from a customers' perspective into the profit optimisation calculus of the network operator. With it, the service quality incentive reflects the costs incurred by customers affected by a poor quality level. Thus, the network operator will aim at providing quality up to the level where the marginal cost of quality equals the reward offered and therefore aims at a socially desirable quality level (Growitsch et al. 2009). This economically efficient approach raises the question how network operators actually respond to the introduction of such regulatory instruments in practice. We are particularly interested in the case of Norway since the Norwegian regulator NVE was the first to incorporate customer valuation of service quality in the regulatory scheme and nowadays features a state-of-the-art approach in this context.

Overall, empirical research on the effectiveness of service quality in distribution networks is rather scarce and findings are heterogeneous. Ter-Martirosyan (2003)

---

[4] For other instruments to regulate the different dimensions of service quality, such as publication of data on company performance, (minimum) quality standards or premium quality contracts, please refer to Fumagalli et al. (2007).

[5] In the remainder of the paper, we focus on ENS since this is the regulatory indicator employed in our case-study Norway.

[6] For further discussion on the choice of the regulated indicator, please refer to Fumagalli et al. (2007).

[7] Some of these countries such as the UK or the Netherlands also employ other instruments of service quality regulation. The Netherlands for instance additionally apply compensation payments in case a predetermined continuity of supply standard is breached, whereas the UK also sets guaranteed standards for commercial quality.

analyses the impact of incentive regulation on the duration and frequency for electric outages for a panel of 78 US utilities. She finds that incentive regulation is associated with an increase of outages. Moreover, the study detects that the number of outages decreases with the introduction of explicit quality benchmarks. Korhonen and Syrjänen (2003) find an improvement in technical efficiency after introducing a continuity of supply indicator (interruption time per customer) in their Data Envelopment Analysis (DEA) of 106 Finish distribution companies. A report by CEPA (2003) applies a two-step DEA model to cross-sectional data for the UK. They find no significant correlation between technical efficiency measures and continuity of supply in terms of customer minutes lost. Ajodhia et al. (2004) apply a DEA and a Corrected Ordinary Least Square (COLS) model to a cross-sectional sample of 44 electric utilities from the UK, the Netherlands, Hungary, and Malaysia, reporting a significant efficiency increase when quality is taken into account, especially for smaller network operators. Giannakis et al. (2005) carried out a DEA-based quality incorporated efficiency study on UK electricity distribution companies. They show that technical efficiency does not necessarily also involve high service quality. Moreover, they find that quality incorporated regulatory benchmarking is superior to cost-only approaches. Jamasb and Söderberg (2009) analyse the effects of the application of norm models within an ex-post incentive regulation of electricity distribution networks in Sweden. In the examination of the companies cost and service quality performance they find that service quality regulation has not affected the relative performance of utilities.

Some recent studies performed by Edvardsen et al. (2006) and Burger and von Geymueller (2007a, b) specifically examine the efficiency of Norwegian distribution networks. Edvardsen et al. focus on the general productivity of the networks. They find a productivity improvement albeit flattening out as from 2000. They generally explain this decrease with the introduction of new regulatory requirements and a potential retention in efficiency awaiting changing regulatory parameters. Burger and von Geymueller (2007a) find that quality regulation induced Norwegian network operators to optimise their quality strategy from a social point of view based on a DEA analysis and Malmquist indices for the period 1999–2005. However, their sample covers a rather limited number of observations (31 distribution companies), which might involve an uncontrolled sample bias. Indeed, in another paper, the authors find that ENS was reduced more significantly prior to the introduction of quality regulation than afterwards (Burger and von Geymueller 2007b).

Albeit previous empirical research addresses service quality and/or productivity and welfare related issues there is—to our best knowledge—no empirical case study that clearly focuses on the impact of WTP-based continuity of supply regulation on the efficiency of distribution networks.

Thus, the aim of this paper is to shed some empirical light on this issue by assessing whether WTP-based[8] service quality regulation has a noticeable effect on

[8] In the remainder of this paper, we use incentive based service quality regulation, WTP-based service quality regulation, and CENS-regulation as synonyms.

the social cost efficiency of distribution networks in Norway and to what extent this regulatory instrument motivates a socially desired quality level. We give empirical evidence by means of a Data Envelopment Analysis and associated tests based on a complete dataset of Norwegian utilities, which was prepared by the Norwegian regulator NVE for the purpose of regulatory benchmarking analysis. The results enable us to discuss the effectiveness of service quality regulation based on customers' WTP and the impact on the quality level in Norway.

The remainder of this paper is as follows: In Sect. 2 we explain the economics of WTP-based service quality regulation. Section 3 describes the Norwegian regulatory approach in terms of service quality regulation and gives empirical evidence. Section 4 concludes and highlights policy implications.

4.2 WTP Based Service Quality Incentives and the Optimal Quality Level

Regulating service quality regulation challenges regulators both methodologically and practically. More specifically, the regulator has difficulties to define and measure precisely the desired service quality level delivered to consumers. One consequence of this lack of verifiability is the difficulty for the regulator to adequately reward or penalise the regulated utility for providing service quality. Therefore, indirect regulatory instruments have to be employed to motivate the desired service quality level (Sappington 2005). Among these instruments arise incentive based continuity of supply schemes based on reliability indicators. The latter are an innovative means to induce regulated companies to deliver the socially desired service quality level.

From a distribution network operator's perspective, providing quality has a certain cost, both in form of capital expenditures (CAPEX) and operation and maintenance expenditures (OPEX). These network cost (together: total expenditures, TOTEX) increase with higher quality levels, marginal improvements being more costly when quality is already high. Likewise, the customers' benefit increase (or cost associated with poor quality decrease) with a higher quality level. From a social perspective, reflecting the optimisation calculus of both customers and companies, the optimal quality level is where the marginal customer benefit of additional quality equals the marginal cost of supplying it. In other words, the optimal quality level corresponds to the minimum of a total cost function incorporating companies' cost in providing quality and the costs experienced by customers due to a poor quality level (Fumagalli et al. 2007).

Taking this theoretical setting for a socially optimal quality level into account, the regulator's task is to include the benefit from service quality (from a customer's point of view) into the firm's decision-making process. In general terms, the customers' preferences for a certain quality level represent their WTP for it. As the quantification of this critical parameter is rather difficult, WTP is usually approximated with its inverse, this is the customers' interruption cost (IC) due to

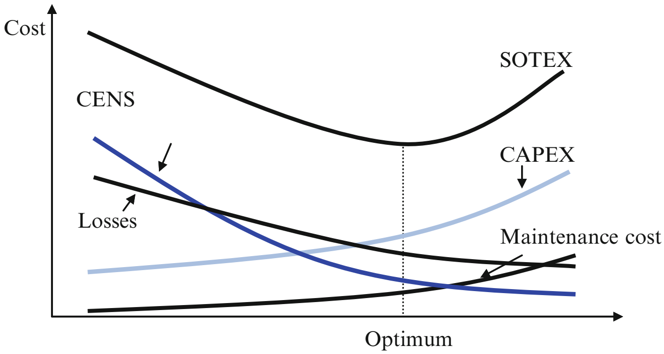


Fig. 4.1 Trade-off between CENS and TOTEX

a poor service quality level (Ajodhia 2006). Service quality incentives for the regulated company are thus calculated by multiplying the respective ENS[9] with the IC for a certain customer group. More specifically, this constitutes the *external* cost of energy not supplied (CENS) which are then internalised in the network operators profit maximisation by regulation.[10] Within such an indirect incentive scheme the regulated firm will aim to optimise its trade-off between CENS and TOTEX. These together form the total social cost (SOTEX) of network provision. The more the network operators invest in network reliability to reduce CENS, the higher TOTEX becomes. At some point, the companies will reach the theoretical optimal quality level where the marginal customer and, by regulation, operator's benefit of additional quality equals the marginal cost of supplying it, in other works where the sum of CENS and TOTEX is lowest (as illustrated in Fig. 4.1). This implies that network operators will only increase quality as long as this leads to a net reduction in SOTEX, or until the marginal costs to provide more quality equal the reduction in CENS incurred by customers (Ajodhia 2006).

Now, these economic considerations imply the necessity to carefully distinguish between a maximum quality level, which constitutes only a unilateral optimisation from either the companies' or the customers' perspective and a socially optimal level for both players. The social optimum, however, is only obtained when SOTEX are taken into account.

[9] Alternatively the SAIDI may be used as regulatory indicator to represent the companies' performance in terms of service quality. A discussion on the difference between ENS and SAIDI can be found at Fumagalli et al. (2007). Within the scope of our paper, we focus on ENS only.

[10] Another regulatory challenge is to adequately approximate the CENS for the regulatory formula with the customer WTP for service quality. A number of regulators have found consumer surveys of WTP for network reliability useful in setting service quality incentives. Different methods can be used to measure WTP. For an overview of the most prominent, please refer to Growitsch et al. (2009).

This section briefly presented the theory of WTP-based service quality incentives applied for indirect regulatory measures to motivate a socially desirable service quality level. In the following section we practically assess the effectiveness of such regulatory instruments by means of a concrete case study.

4.3 The Norwegian Example

4.3.1 Overview

This section explores and assesses the development path of quality regulation in Norway, one of the pioneering countries in this field. The objective is to further scrutinise the issue of implementing quality incentives based on customer WTP for continuity of supply, and to analyse the impact of such regulatory measures on the efficiency of the Norwegian network operators by means of a concrete case study. After a brief description of Norwegian quality regulation, we analyse the adaptation of the network operators in terms of their improvement in social cost efficiency. Comparing this with the development of private cost (TOTEX) efficiency provides evidence of the effectiveness of quality regulation in Norway.

4.3.2 Quality Regulation in Norway: Development and Status Quo

The first features of quality regulation were introduced after regulatory reform in 1991 by the Norwegian regulator (NVE). In 1995, NVE implemented a standardized reporting system for interruptions and outages called Fault and Supply Interruption and Information Tool (FASIT). As a result, network operators were obliged to report all interruptions and outages longer than 3 min (Brekke 2007). From 1997 on, network operators at 33–420 kV were required to report any incidents, disturbances and system failures. Simultaneously, a revenue cap was introduced yet without any incentive for quality management, thus leading to a tendency towards underinvestment. Likewise standardised methods to compute the ENS per customer category were set up and a reporting system was made mandatory. Eventually in 2001, a quality term based on the CENS was incorporated into the regulatory formula to determine the revenue cap for the second regulatory period (2001–2006).

The main objective of the CENS-arrangement is to give the network owners incentives to plan, operate and maintain their networks in a socio-economic optimal way and thereby provide a socio-economic optimal level of continuity of supply. In more general terms, the objective of service quality regulation in Norway is to obtain a service quality level that is beneficial for the society as a whole. This does not necessarily include the need for a general improvement in the current service quality level, but rather the objective to create a social awareness for service quality defining what the socially desirable actually looks like (Brekke 2007).

As a consequence, the utilities revenue cap was adjusted in accordance with the customers' interruption cost during the regulatory period 2001–2006. In pursuing this approach all planned and unplanned interruptions longer than 3 min in networks over 1 kV were considered. Based on estimates of expected ENS and average outage costs per customer group, the underlying model annually computes the expected outage costs per network operator. The latter particularly depends on two determinants: the customer group and the type of interruption (planned or unplanned) as illustrated by (4.1):

$$IC = \sum_{n,m} ENS_{n,m} \cdot c_{n,m} \tag{4.1}$$

where

IC = Cost of energy not supplied/Outage cost (€)
ENS = Energy not supplied (kWh)
c = average specific outage costs
n = customer group
m = planned, unplanned interruption[11]

ENS is defined as the amount of energy that would have been supplied to the customer if there had been no interruption. This amount can be estimated by means of FASIT, which provides a uniform standardised methodology. The average specific outage cost (c) can however be appraised based on customer surveys that have been conducted since 1991 (Langset et al. 2001). Table 4.1 illustrates the respective values per customer group resulting from a nationwide survey conducted in 2002.[12]

Table 4.1 Specific outage costs in the Norwegian CENS arrangement 2003–2006 (€/kWh)

Customer group	Notified outage costs	Non notified outage costs
Industry	8.25	5.75
Trade and services	12.38	8.50
Agriculture	1.88	1.25
Households	1.00	0.88
Public facilities	1.63	1.25
Wood processing	1.63	1.38

Source: Brekke (2007), own translation

[11] Different incentive rates are used for notified and non-notified interrupstions (see Table 4.1).

[12] Recently the Norwegian regulator NVE has conducted a new survey on consumer valuation of interruptions and voltage problems (Kjølle et al. 2008). The survey finds a significant increase in the customers' costs since the 1991 survey for all customer groups and particularly for the agricultural group. Amongst others, the newly identified CENS for short interruptions ≤3 min will be incorporated in the CENS-arrangement as from 2009 (see footnote 13).

The outage cost in the Norwegian CENS-arrangement comprises both notified and non-notified interruptions, featuring different outage costs for those different types of interruptions.

Network operators are also set individual quality targets. In other words, the outage costs for all customers that are connected to the distribution networks are capped at a specific sum. To this end the expected value for ENS for each network is estimated by means of regression analysis (4.2). This analysis uses parameters such as network structure, number of transformers, geographic and climatic factors. Panel data from previous years provide the historical values for ENS. Consequently, quality targets can be derived from the expected value of outage costs.

$$E(IC) = \sum\nolimits_{n,m} E(ENS)_{n,m} \cdot c_{n,m} \tag{4.2}$$

where

E (IC) = Expected outage costs [NOK]
E (ENS) = Expected ENS [kWh]

$C_{n,m}$ as above.

At the end of the year the difference between expected and actual outage costs is calculated. In the case of a positive difference, i.e., the reliability is higher than expected, the difference is added to the revenue cap. In the case of a negative difference, the amount is subtracted from the revenue cap. This mechanism is illustrated by (4.3) and Fig. 4.2.

$$dR = E(IC) - IC \tag{4.3}$$

where

dR = change in Revenue Cap

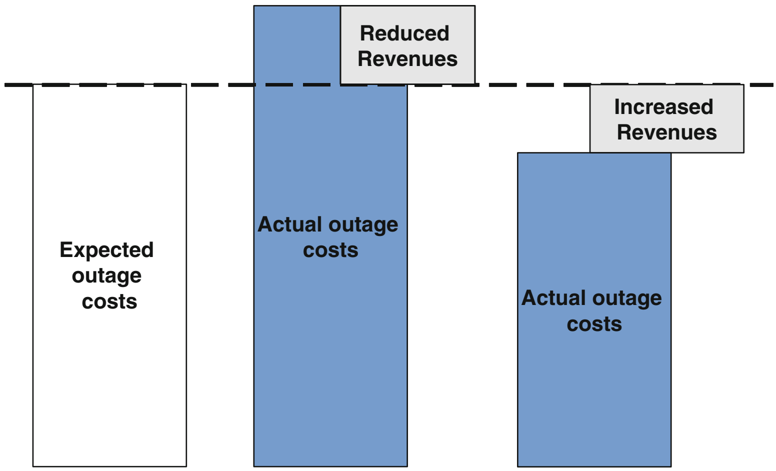


Fig. 4.2 Outage costs and revenues. Source: Brekke (2007)

The calculations described above are carried out 1 year after the determination of network charges by the network operator. Therefore a gap usually occurs between the expected (allowed) revenues and the actual revenues as already illustrated in Fig. 4.2. If the difference is to the benefit of the network operator, the firm is obliged to pay back the windfall profit through lower network charges to its customers in the following years. Conversely the firm is allowed to be compensated for a potential loss through higher network charges. Thus an increase in reliability (i.e., a decrease in outage costs IC) leads to higher revenues whilst a decrease in quality leads to lower revenues. Given this mechanism (4.4) applies:

$$\mathrm{R'} = \mathrm{IC'} \tag{4.4}$$

where

R' = marginal revenue
IC' = marginal outage costs for a specific customer group

Moreover, the economic costs for network operation can be considered as the result of company specific capital expenditures (CAPEX) and operational expenditures (OPEX) as well as the outage cost of the customers as shown under (4.5).

$$\mathrm{C = OPEX + CAPEX + IC} \tag{4.5}$$

The economic optimum for marginal outage costs results from a minimisation of (4.6), given that

$$\mathrm{OPEX' + CAPEX'} = \mathrm{IC'} \tag{4.6}$$

Consequently the profit of a network operator can be expressed as

$$\Pi = \mathrm{R - OPEX - CAPEX} \tag{4.7}$$

Therefore a profit-maximising network operator would act on the assumption that

$$\mathrm{OPEX' + CAPEX' = R'} \tag{4.8}$$

Taking these assumptions into account as per (4.4), (4.6) and (4.8) we deduce that a profit-maximising network operator under the Norwegian regulatory regime would strive towards a social optimum by minimising overall economic (social) costs.

Brekke (2007) concludes that the implementation of the quality regulation system has sensitised the network operators to outage costs incurred by their customers. This motivated a change in the operation and management of their assets. Moreover, the regulatory regime allows for a clear definition of

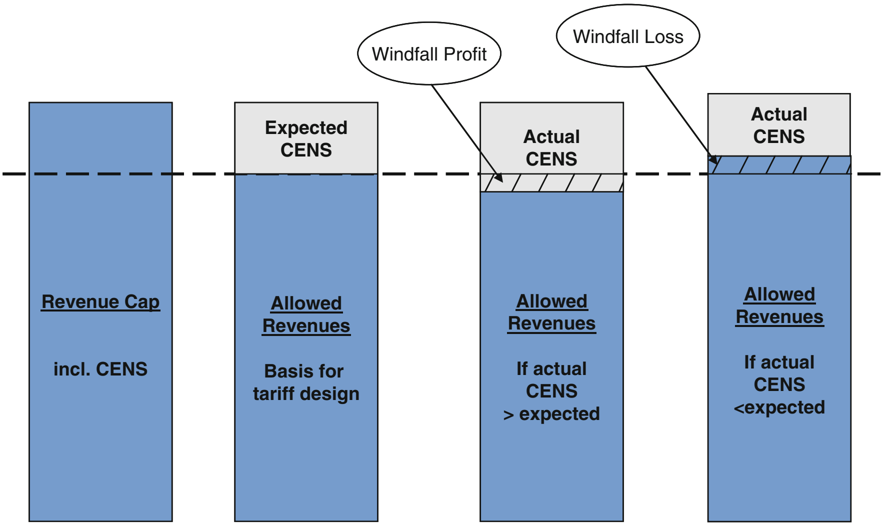


Fig. 4.3 Revenue Cap and outage costs since 2007. Source: Brekke (2007)

responsibilities in the network and therefore, facilitates performance improvements. Brekke detected, however, some shortcomings in the system such as the unsatisfactory recovery time following an interruption for those clients for whom the CENS-regulation does not set strong enough incentives. Moreover, short interruptions are not taken into account, which may lead to higher costs to the affected customers.

The shortcomings detected by Brekke (2007) have partly been addressed by amendments to the regulatory regime with the start of the new regulatory period in 2007. Since then outage costs have been directly integrated into the calculation of the revenue cap (Fig. 4.3). Thus, the costs incurred for the provision of a certain quality level are considered as part of OPEX and feed into the DEA-based benchmarking; and the revenue caps are adjusted on an annual basis with respect to the actual level of interruptions.

In parallel, another component of quality regulation has been introduced, namely direct compensation payments. As a result, network operators are obliged to pay direct compensation to those customers affected by interruptions longer than 12 h (Brekke 2007).

These payment obligations follow the schedule below:

- For 12–24 h: 600 NOK (approx. 70 €).
- For more than 24 till 48 h: 1.400 NOK (approx. 160 €).
- For more than 48 till 72 h: 2.700 NOK (approx. 310 €).

Additionally, 1.300 NOK (approx. 150 €) applies to each subsequent 24-h period. However, the payments should not exceed the annual tariff payments.

Moreover, short interruptions lasting from 1 to 3 min as well as the time dependency are integrated into the CENS-system as from 2009.[13]

The previous sub-section provided an overview of the evolution of quality of supply regulation in Norway. In summary we conclude that Norway has a mature system for determining the external costs of quality and for incorporating them into the regulatory formula.

It is also worthwhile to look behind the scenes of the Norwegian system in order to gain empirical evidence of the actual impact of service quality regulation on the efficiency situation of Norwegian network operators and the service quality level as a whole. This review is carried out in the following section.

4.3.3 *Data and Methodology*

In order to examine the performance of the Norwegian approach to service quality regulation, we use a panel dataset for 131 Norwegian distribution network operators (DNO) from the period 2001–2004.[14] In Norway, the DNOs are mainly publicly owned. The Norwegian Energy Act stipulates full ownership unbundling between generation and supply functions on the one hand and transmission and distribution on the other.[15]

The productivity analysis method used is Data Envelopment Analysis (DEA) technique.[16] DEA measures the relative efficiency of a company relative to the best performing companies (peers) by means of a non-parametric, linear frontier over the sample. This piece-wise approach aims at fitting a linear "hull" around the data assuming that this hull adequately forms the frontier of the most productive firms by

[13] Kjølle et al. (2009) describe the latest changes to service quality regulation in Norway. The motivation for this was to achieve the most optimal level of continuity of supply for the society as a whole. Given the fact that recent customer surveys in the Norwegian electricity distribution sector found that annual costs for short interruptions ≤ 3 min were associated with the same WTP as for long interruptions >3 min, short interruptions are incorporated in the CENS-arrangement as from 2009. Moreover, customer surveys indicated that time dependency in interruption costs was found to be significant. Therefore, cost functions are corrected by monthly, weekly and daily variations in order to stimulate more cost-efficient maintenance activities, this is e.g. in periods where the interruption cost is low. However, the temporal value of CENS is not taken into account in our analysis since the related regulatory instruments only became effective in 2009, whilst our data sample ends with the year 2004.

[14] For the following discussion, it should be noted that the time horizon of the analysed data ends at 2004. Hence, the companies within our sample could not react to the latest features of quality regulation that were introduced in the second regulatory period, this is basically the mandatory reporting of interruptions and the introduction of compensation payments The enhancements to the regulatory model as discussed in Sect. 3.2 cannot be tested.

[15] The Norwegian TSO Statnett SF is not part of any vertically integrated undertaking. With regard to our analysis, transmission data in any respect is excluded since it is out of the control of a single DNO, therefore.

[16] For a detailed introduction see Coelli et al. (2005) and Greene (2007).

means of a deterministic approach with multiple inputs and outputs. The resulting efficiency score reflects the amount by which a given company could improve its productivity relative to its peers. The most efficient company is assigned an efficiency score of one given that it scores best by minimising its inputs for a given level of output. This approach has been first proposed by Farrell in 1957. In the following two decades options for mathematical programming have been suggested by Boles (1966), Shephard (1970) and Afriat (1972). Nevertheless, only in the late seventies this method eventually attracted global interest following a paper published by Charnes et al. (1978) which first introduced the term "DEA". In this paper the authors argued that the main task of DEA was to compute an efficiency score for each company based on a linear program formulation. Moreover Charnes et al. (1978) advocated an input-oriented model assuming constant returns to scale (CRS) to estimate the production frontier. The CRS approach assumes that changes in output result from proportionate changes in input. Firms having altered their sizes and thus diverging from the optimal operating scale will therefore be detected as inefficient by the model. Alternatively the assumption of variable returns to scale (VRS) provides for a limited control of the firms on their scale and therefore implies a correction of the efficiency score for such differences. Consequently, only firms operating on a similar scale will be compared to each other (Ajodhia 2006). In the following example, we assume constant returns to scale (CRS) since the networks operators may, in general, be able to optimize their size and scale. By that, we consider inefficiencies due to a deviation from optimal scale.[17] A CRS input-oriented frontier is calculated by solving the linear optimization program in (4.9) for each of N companies. Moreover, it is assumed that the companies use K inputs and M outputs (Shephard 1970):

$$
\begin{aligned}
&\max \quad \theta, \\
&\text{s.t.} \quad -y_i + Y\lambda \geq 0, \\
&x_i/\theta - X\lambda \geq 0, \\
&\lambda \geq 0,
\end{aligned}
\tag{4.9}
$$

where X is the K*N matrix of inputs and Y is the M*N matrix of outputs. The i-th company's input and output vectors are represented by x_i and y_i respectively. λ is a N 1 vector of constants and θ is the input distance measure.

We chose to use the DEA technique as it is particularly suitable for multiple input and multiple output efficiency analysis and for this reason it is often the method of choice by most regulators that practice benchmarking including the Norwegian regulator (see Jamasb and Pollitt 2001).

A in its original form, however, is unable to provide unbiased efficiency estimates and confidence limits for the efficiency scores. The theoretical bias is evident since the observed input–output combination is just a fraction of any possible one:

[17] NVE uses a variable returns to scale (VRS) model. However, Kittelsen (1994) suggested to use CRS in order to encourage cost saving restructuring also in terms of network size.

$(\mathrm{x}, \mathrm{y}) \subseteq (X, Y)$. This implies that the estimated production set $\hat{\psi}$ is a just a subset of $\Psi, \hat{\psi} \subseteq \psi$. Efficiency is estimated and compared within a restricted sample and the estimator is upward biased as a result. We apply a bootstrap procedure suggested by Simar and Wilson (1998) to overcome this problem. It provides an estimate for DEA's upward efficiency bias and confidence intervals by drawing random samples from the efficiency scores' truncated probability density functions.

As highlighted above, DEA determines the efficiency score of a firm compared to its peers and therefore, indicates the catch-up potential within a given sample. For the purpose of this paper the cost of service quality is incorporated into the benchmarking. Therefore, it is crucial to provide for the ambivalent relationship between productive efficiency and quality. In general one may assume that higher quality levels lead to higher costs. In a cost-based DEA, companies operating at higher quality levels would therefore likely score worse than their efficiency-oriented counterparts albeit running their business to the benefit of quality. This potential trade-off can be reduced by incorporating SOTEX into the DEA and thus accounting for the provision of quality (Ajodhia 2006).

The model specification incorporates total expenditures TOTEX and SOTEX, respectively. These are considered separately as a single input in monetary terms. Hence, we use two models, one with TOTEX and the other one with SOTEX as input variable. In Model one TOTEX describes the sum of OPEX and CAPEX, both influencing the productivity of the network operator without explicitly considering quality aspects. By contrast, Model two incorporates SOTEX as the input variable in order to reflect the impact of quality incentives. SOTEX is the sum of TOTEX (corporate production costs) and the external costs of low quality, i.e., the CENS incurred by customers. Thus, the resulting efficiency scores of SOTEX reflect the ability of the network operator to balance the trade-off between efficient costs and quality (Ajodhia 2006).

We use a simple model with one input and two outputs. Inputs are either TOTEX or SOTEX. The outputs consist of energy supplied and the number of customers, following the most frequently used output variables in international regulatory benchmarking (Jamasb and Pollitt 2001).[18] Although the two cost drivers form one joint service in electricity distribution they are considered separately since they drive different cost categories, namely fixed and variable costs (Growitsch et al. 2009). The model assumes input-orientation, i.e., the efficiency score depends on the ability of the network operator to minimise its inputs given a fixed vector of outputs. Table 4.2 shows the descriptive statistics of our sample aggregated for the considered period and individually for the respective years. Table 4.3 exhibits the mean for the years 2001–2004.[19]

[18] NVE's original model is a little more complex, dividing the input side into different kinds of cost (wages, other OPEX, network losses and CAPEX). For analytical reasons, we have combined the various inputs to a single private cost input (TOTEX), as we are not focusing on optimal factor allocation within the firm, but of private and social cost efficiency.

[19] For an overview of the descriptive statistics per year, see Appendix.

Table 4.2 Descriptive statistics of the sample (aggregated)

Variable	Mean	Std. Deviation	Minimum	Maximum	Cases
SOTEX (k€)	76,406	166,517	2074	1,598,890	524
TOTEX (k€)	74,067	161,395	2074	1,561,140	524
Final customers (no.)	19,784	52,854	429	516,339	524
Energy supplied (MWh)	523,231	1,481,630	7470	15,482,400	524

Table 4.3 Means for the period 2001–2004

Variable	Mean 2001	Mean 2002	Mean 2003	Mean 2004
SOTEX (k€)	77,830	79,224	76,646	75,857
TOTEX (k€)	75,783	77,372	73,396	73,510
Quality cost	2047	1852	3249	2348
Final customers (no.)	19,912	19,956	20,083	20,216
Energy supplied (MWh)	559,071	540,384	501,420	520,255

With regard to SOTEX we find that costs slightly increase in 2002 followed by a decline in the following years. A similar development can be observed for TOTEX. Accordingly the cost of quality decreases in 2002 followed by a significant increase in 2003. Simultaneously the standard deviation and the maximum more than double compared to 2002. This development suggests that a significant event took place in 2003 featuring increased prices. Looking at the output variables, the final customers slightly increase after an initial stagnation, whilst the energy supply declines over the period. Overall we show in Table 4.3 that there is only a rather marginal gap between TOTEX and SOTEX. Moreover, homogenous trends can be reported for SOTEX and TOTEX.

Based on this first impression, we hypothesize that the external costs of quality have a small effect on the cost and, as a result, the incentives of the Norwegian network operators. In the following section we test this hypothesis by analysing the results of the DEA regarding the efficiency of the sample of Norwegian network operators.

4.3.4 Estimation and Results

Table 4.4 shows the bootstrap results of the DEA for Model one (input: TOTEX) and Model two (input: SOTEX). As expected, the unbiased estimates obtained from bootstrapping are slightly lower than the original scores. In order to test whether the annual average unbiased efficiency scores for TOTEX and SOTEX differ significantly from each other, we use the non-parametric Wilcoxon ranksum test.[20]

[20] The Wilcoxon ranksum test, also Mann-Whitney-U-Test, is a non-parametric test that analyses whether two independent groups belong to the same population (see Cooper et al. 2006)

Table 4.4 Efficiency for model 1 (TOTEX)

	Model 1 (TOTEX)			Model 2 (SOTEX)		
Variable/ Year	Mean (%)	Mean[a] (unbiased) (%)	Std. Deviation (%)	Mean (%)	Mean[a] (unbiased) (%)	Std. Deviation (%)
2001	62.76	60.97	14.71	62.12	60.33[b]	14.68
2002	58.15	55.81[c]	15.50	58.91	56.64[b]	15.60
2003	56.45	53.58	14.36	56.51	53.82	14.97
2004	57.31	54.22	14.25	57.81	55.16[b]	14.65

[a]Efficiency score bias corrected via bootstrap (100 replications)
[b]Indicates a difference to TOTEX on 5 % level of significance
[c]Indicates a difference to prior year on 5 % level of significance

Comparing the results for TOTEX and SOTEX for each year, we find no systematic differences between the models.[21]

Analysing the development of TOTEX estimates only, the efficiency decreases significantly after the first year and remains statistically constant from 2002 to 2004.[22]

Concerning the SOTEX cross-sections, the efficiency scores do not vary from year to year. Comparing average efficiencies from 2001 to 2004, however, indicates marginally but statistically significantly lower social cost efficiency 4 years after the introduction of the CENS regulation.

Overall we find that TOTEX and SOTEX almost develop in similar manners, corroborating the initial hypothesis we made. Moreover, the Wilcoxon ranksum test showed that there is no significant difference in the unbiased efficiency score between the years 2002 and 2004, neither for TOTEX nor SOTEX. The reduction in TOTEX and SOTEX efficiency in 2004 relative to 2001 coincides with the development of the average efficiency score shown in the descriptive statistics.

A closer examination of efficiency scores on a per company basis, however, shows that the efficiency scores for individual firms can change significantly from year to year. At the same time, the TOTEX and SOTEX scores, for a given year, are rather similar. Figures 4.4 and 4.5 show the utilities' efficiency scores (Y-Axis) for 2001 in increasing order relative to those of 2002–2004 (Company ID, X-Axis). Moreover, the figures show that the scores of more efficient utilities in 2001 (i.e., right hand side of the figures) also tend to be higher than in subsequent years. However, the peers change over time.

The analysis of the efficiency development shows that the introduction of quality regulation did not significantly change the efficiency scores of the companies.

[21] In the year 2001 TOTEX scores are marginally but significantly higher, in the years 2002 and 2004 significantly but marginally lower than the SOTEX scores. In 2003, there are no significant differences.

[22] To control for possible scale effects, we also calculated the annual efficiency averages under VRS. The results differ in levels, but not in their economic interpretation. Detailed information may be obtained from the authors upon request.

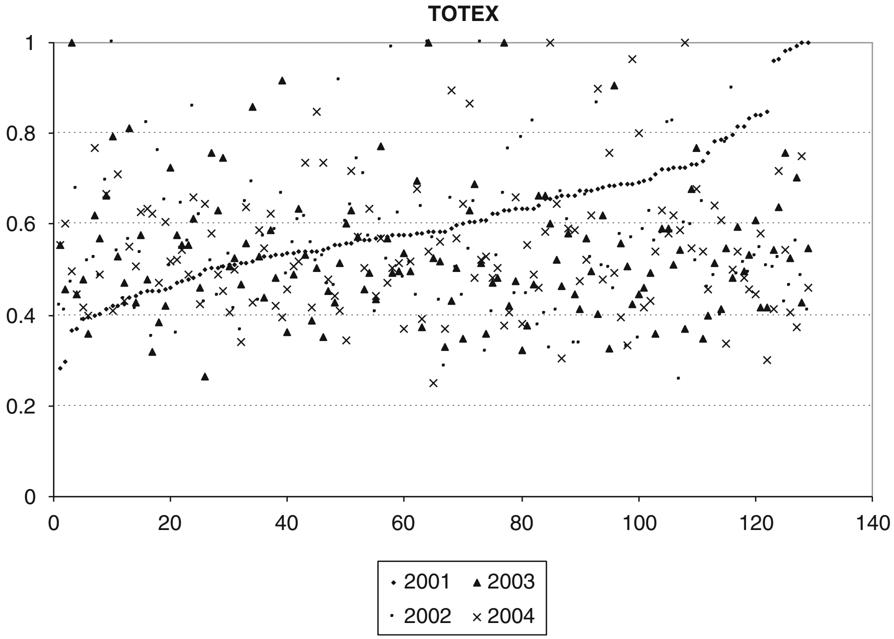


Fig. 4.4 TOTEX efficiency scores by company and year

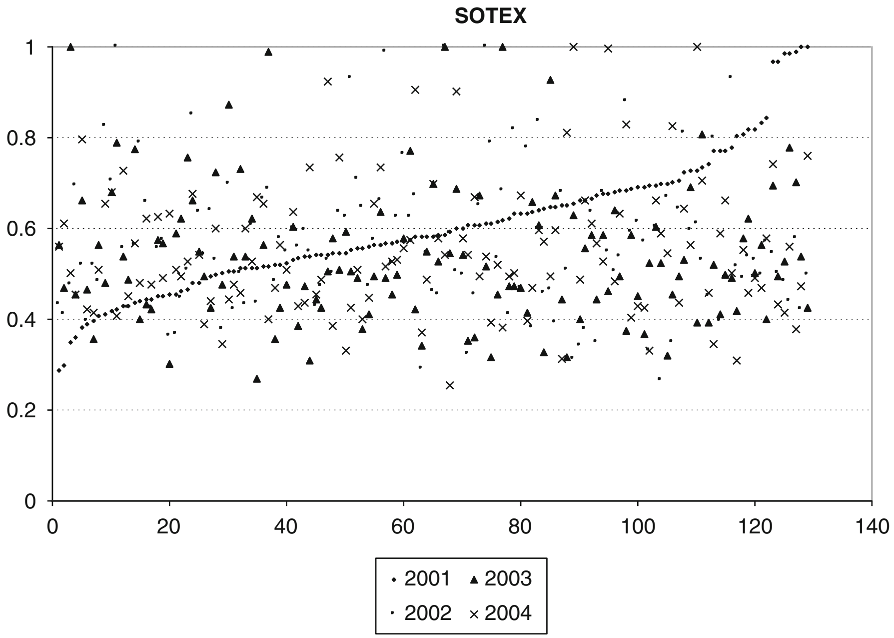


Fig. 4.5 SOTEX efficiency scores by company and year

Moreover, it appears that the external costs for quality are quite low, which is proven by the fact that the difference between TOTEX and SOTEX is nearly zero. These findings are substantiated by the fact that the costs of energy not supplied in Norway are rather low compared to other European countries (Ajodhia 2006 and Fumagalli et al. 2007).

The results suggest that the introduction of quality regulation in Norway does not have a strong negative impact nor does it economically conflict with cost efficiency of the networks—i.e., the external quality costs play a relatively minor role. Moreover, the level of quality appears to be close to the social optimum, which explains the limited impact of CENS-regulation on the efficiency scores. However, benchmarking results in general and the empirical findings for the Norwegian example in particular have to be treated prudently since they only provide a first quantitative approximation of the implications of service quality regulation.

4.4 Conclusions

The objective of this paper was to scrutinise the issue of quality-related incentives based on customers' WTP for continuity of supply regulation and to analyse the impacts of such indirect regulatory measures by means of a concrete case-study.

After a theoretical introduction on WTP-based service quality regulation and the optimal quality level we described how Norway, a pioneer in this field, has put this regulatory approach into practice. In summary we conclude that Norway has a mature system for determining the external costs of quality and for incorporating them into the regulatory formula. In the following, we empirically examined how the distribution network operators adapted to the Norwegian CENS-arrangement. In order to do this, we analysed whether the distribution network operators changed their quality-related optimisation strategies reflected by their cost efficiency developments. The results indicate that incorporating the external cost of quality in incentive regulation benchmarking models has not played a major role during the period 2001–2004.

A first intuitive explanation for this may be that the service quality level was already close to the social optimum prior to the implementation of quality regulation. This confirms the statement by Brekke (2007) that the actual level of service quality was generally perceived as "satisfactory" during our observation time. This rationale should however, be pulled together with the current enhancement to the CENS-regulation in Norway. The current amendments (incorporation of short interruptions and time dependency in the CENS-arrangement) in Norway are motivated with the need to find the "most optimal" level for the society as a whole (Kjølle et al. 2009). This implies that politics, regulators and society still call for an optimisation potential with regard to the Norwegian service quality level from am social point of view.

As regards the empirical visibility of such endeavours, our quantitative results should, however, be treated with caution since our data panel only covers the period from 2001 to 2004. Moreover, we only focused on TOTEX and SOTEX efficiencies and did not further elaborate on productivity developments and welfare implications due to limited data availability. This caveat indicates that data availability (especially for a longer time horizon) and robustness are limiting factors for this kind of analysis. Moreover, there is a time lag between the introduction of quality regulation and its impact on the investment decisions of network operators. Thus, the full impact of quality and asset management related strategies of network operators might not yet be reflected in the efficiency scores within the time horizon considered in the this study. Nevertheless, our findings generally confirm that the service quality level has already reached an economically reasonable level. The verification whether the optimisation potential towards the most optimal quality level for the Norwegian society as a whole is actually exhausted still requires empirical evidence.

Therefore, future research should in particular address the issue of delayed reactions in terms of continuity of supply improvement. Moreover a parallel analysis of productivity developments based on data reflecting the latest amendments to CENS-regulation should be carried out in order to empirically disclose the step from a *satisfactory* towards a *most optimal* service quality level.

Appendix

See Tables 4.5, 4.6, 4.7 and 4.8.

Table 4.5 Descriptive statistics year 2001

Variable	Std. Deviation	Minimum	Maximum	Cases
SOTEX (k€)	174,288	5045	1,561,070	129
TOTEX (k€)	170,237	4949	1,525,533	129
Quality cost	4384	22	35,537	129
Final customers (no.)	53,461	936	516,339	129
Energy supplied (MWh)	1,571,051	18,720	15,500,000	129

Table 4.6 Descriptive statistics year 2002

Variable	Std. Deviation	Minimum	Maximum	Cases
SOTEX (k€)	177,614	5153	1,598,891	129
TOTEX (k€)	173,678	5054	1,561,144	129
Quality cost	4198	27	37,747	129
Final customers (no.)	53,073	925	508,393	129
Energy supplied (MWh)	1,525,085	17,557	15,000,000	129

Table 4.7 Descriptive statistics year 2003

Variable	Std. Deviation	Minimum	Maximum	Cases
SOTEX (k€)	161,219	5574	1,361,567	129
TOTEX (k€)	153,243	5385	1,273,104	129
Quality cost	8847	39	88,463	129
Final customers (no.)	53,298	927	511,374	129
Energy supplied (MWh)	1,420,952	16,708	14,100,000	129

Table 4.8 Descriptive statistics year 2004

Variable	Std. Deviation	Minimum	Maximum	Cases
SOTEX (k€)	158,467	5807	1,356,415	129
TOTEX (k€)	153,452	5798	1,307,400	129
Quality cost	5507	9	49,015	129
Final customers (no.)	53,671	969	515,152	129
Energy supplied (MWh)	1,463,252	16,504	14,400,000	129

References

Afriat SN (1972) Efficiency estimation of production functions. Int Econ Rev 13:568–598

Ajodhia VS (2006) Regulating beyond price—integrated price-quality regulation for electricity distribution networks. TU Delft/KEMA

Ajodhia VS, Petrov K, Scarsi GC (2004) Economic benchmarking and its applications. In: Proceedings of the 3rd international infrastructure conference on applied infrastructure research, 9 Oct 2004, Berlin

Boles JN (1966) Efficiency squared—efficient computation of efficiency. In: Proceedings of the 39th meeting of Western Farm Economic Association, pp. 137–142

Brekke K (2007) Reliability of supply regulation in Norway, Norwegian Water Resources and Energy Directorate. Energy and Regulation Department, Grid Section

Burger A, von Geymueller P (2007a) Did quality regulation improve welfare in Norway? Combining DEA-analysis and a welfare-framework, Research Institute for Regulatory Economics Working Paper no. 4, Vienna University of Economics and Business Administration

Burger A, von Geymueller P (2007b) Assessing the effects of quality regulation in Norway with a quality regulated version of dynamic DEA. Research Institute for Regulatory Economics Working Paper no. 5. Vienna University of Economics and Business Administration

CEER (2008) 4th benchmarking report on quality of electricity supply 2008. Council of European Energy Regulators, Ref: C08-EQS–24-04, December, Brussels

CEPA (2003) Background to work on assessing efficiency for the 2005 distribution price control review, scope study—final report prepared for OFGEM

Charnes A, Cooper W, Rhodes E (1978) Measuring the efficiency of decision making units. Eur J Oper Res 2(6):429–440

Coelli TJ, Prasada Rao DS, O'Donnell CJ, Battese GE (2005) An introduction to efficiency and productivity analysis, 2nd edn. Springer, New York

Cooper WW, Seiford LM, Tone K (2006) Introduction to data envelopment analysis and its uses, with DEA-solver software and references. Springer, New York

Edvardsen DF, Forsund FR, Hansen W, Kittelsen SAC, Neurauter T (2006) Productivity and regulation reform of Norwegian electricity distribution utilities. In: Coelli T, Lawrence D (eds) Performance measurement and regulation of network utilities. Edward Elgar, Cheltenham

Fumagalli E, Lo Schiavo L, Delestre F (2007) Service quality regulation in electricity distribution and retail. Springer, New York

Giannakis D, Jamasb T, Pollitt M (2005) Benchmarking and incentive regulation of quality of service: an application to the EUK electricity distribution utilities. Energ Policy 33(17):2256–2271

Greene WH (2007) LIMDEP Version 9.0, econometric modelling guide, vol. 1

Growitsch C, Jamasb T, Pollitt M (2009) Quality of service, efficiency, and scale in network industries: an analysis of European electricity distribution. Appl Econ 41(20):2555–2570

Jamasb T, Pollitt M (2001) Benchmarking and regulation: international electricity experience. Util Policy 9(3):107–130

Jamasb T, Söderberg M (2009) Yardstick and ex-post regulation by norm model: empirical equivalence, pricing effect, and performance in Sweden. In: Faculty of Economics, University of Cambridge in its series. Cambridge Working Papers in Economics, no. 0908

Kittelsen S (1994) Effektivitet og regulering i norsk elektrisitetsdistribusjon, SNFrapport 3/94 [in Norwegian]

Kjølle G, Samdal K, Brekke K (2009) Incorporating short interruptions and time dependency of interruption costs in continuity of supply regulation. CIRED 2009 Paper no. 0494. In: 20th international conference on electricity distribution, Prague, June 2009

Kjølle G, Samdal K, Balbir S, Kvitastein OA (2008) Customer costs related to interruptions and voltage problems: methodology and results. IEEE Trans Power Syst 23(3):1030–1038

Korhonen P, Syrjänen M (2003) Evaluation of cost efficiency in Finish electricity distribution. Ann Oper Res 121(1):105–122

Langset T, Trengereid F, Samdal K, Heggset J (2001) Quality dependent revenue caps—a model for quality of supply regulation, Electricity Distribution, Part 1: Contributions. CIRED (IEE Conf. Publication no. 482), vol. 6, S. 5ff

Sappington D (2005) Regulating service quality: a survey. J Regul Econ 27:123–154

Shephard RW (1970) Theory of cost and production functions. Princeton University Press, Princeton

Simar L, Wilson P (1998) Sensitivity analysis of efficiency scores. How to bootstrap in nonparametric frontier models. Manage Sci 44:49–61

Ter-Martirosyan A (2003) The effectives of incentive regulation on quality of service in electricity markets. Department of Economics, George Washington University, Working Paper

Chapter 5
DEA Applications to Major League Baseball: Evaluating Manager and Team Efficiencies

Brian D. Volz

Abstract This chapter provides a brief review of previous applications of data envelopment analysis to the professional baseball industry followed by two detailed applications to Major League Baseball. The first application presents a DEA model which evaluates the efficiency of Major League Baseball managers. This application calculates the output oriented technical efficiencies and super efficiencies for managers from 1986 to 2011. The model assumes that managers are given a certain set of players and evaluates how well they turn those players into wins. A second DEA model is used to evaluate the allocation of resources within a team and identify opportunities for improvement. This model assumes that teams are constrained by their total player budget and evaluates how well the team produces both wins and playoff wins given that budget. A ranking of franchises based on their efficiency from 1986 to 2011 is presented. A method for evaluating the allocation of resources within a specific team during a specific season is also described. Taken together these two applications should provide the reader with a basic understanding of the DEA methodology and how it can be applied to the professional baseball industry.

Keywords DEA • MLB • Baseball • Managers • Teams • Efficiency

5.1 Introduction

There are few industries in which firms keep as detailed records of their inputs and outputs as professional baseball. Detailed and comprehensive statistics on the performance of teams and individual players can be found for seasons dating back to the 1870s. These statistics can range from as general as who won a particular game to as specific as what type of pitches were thrown to a particular batter during a particular inning of a particular game. This obsessive record keeping

B.D. Volz (✉)
Department of Economics and Global Studies, Assumption College, Worcester, MA, USA
e-mail: bd.volz@assumption.edu

J. Zhu (ed.), *Data Envelopment Analysis*, International Series in Operations Research & Management Science 238, DOI 10.1007/978-1-4899-7684-0_5

has made professional baseball a popular subject amongst economists. There exists a large sports economics literature, much of which is dedicated to the sport of baseball. This literature is much too large to be summarized in any one chapter and therefore only those academic papers which employ the methodology of data envelopment analysis (DEA) are mentioned here.

DEA is particularly applicable to professional baseball in that the inputs and outputs of specific players and teams can be accurately defined and calculated. Economists have taken advantage of this characteristic and applied data envelopment analysis to a range of topics within professional baseball over the past two decades. Several applications, such as Mazur (1994) and Anderson and Sharp (1997) use DEA to evaluate the performance of individual players. Sueyoshi et al. (1999) also evaluate the performance of individual players using what they refer to as a Slack-Adjusted DEA model. Howard and Miller (1993) propose a DEA model in which they use performance statistics as inputs and salaries as outputs. This model enables them to identify players who are under or overpaid. In another application related to player performance, Chen and Johnson (2010) use DEA to show that the criteria for evaluating pitchers have changed over time. Miceli and Volz (2012) also examine individual players in their analysis of hall of fame voting. They employ a DEA model to calculate the maximum hall of fame votes players could have produced given observed voting behavior. Their results suggest that up to a third of the hall of fame could be replaced with more deserving players if every player received their maximum votes.

Numerous other studies have chosen to focus on teams rather than individual players. These studies generally include measures of player performance or salaries as inputs and wins as outputs. Einolf (2004) uses player salaries as inputs to a DEA model in order to evaluate the efficiency of baseball and football teams. He finds that large market baseball teams are more likely to overspend on players than small market teams. Sexton and Lewis (2003) also include player salaries as inputs when analyzing the efficiency of Major League teams. However, they utilize a two-stage DEA model in order to distinguish between front office and on-field inefficiencies. Lewis and Sexton (2004a) also use a two-stage DEA model of Major League teams to illustrate a method for incorporating negative inputs and outputs into DEA models. Lewis et al. (2007), using another two-stage DEA model, examine the minimum salary necessary for teams to be competitive in Major League Baseball and trends in that salary over time. They, like Einolf (2004), find that large market teams are more likely to overspend on players. Additionally, Lewis et al. (2009) use a network DEA model, developed in Lewis and Sexton (2004b), in their evaluation of the relative importance of efficiency and capability in team performance. They find that efficiency does contribute to regular season performance but is less important than capability. Volz (2009) also uses player salaries as inputs to the production of wins in order to evaluate the performance of Major League managers. The resulting efficiency scores are then included as a covariate in determining the effect of race on the survival of Major League managers.

A version of Volz's DEA model which has been updated and expanded to include super-efficiency is presented in detail in this chapter. A second DEA model which can be used to evaluate the allocation of resources within a team and identify opportunities for improvement is also presented. Taken together these two applications should provide the reader with a basic understanding of the DEA methodology and how it can be applied to the professional sports industry. The DEA techniques employed in these applications are based on the methods introduced by Charnes et al. (1978). The specific models used are a variable returns to scale model similar to that of Banker et al. (1984) and a super efficiency model similar to that proposed by Andersen and Petersen (1993). The specifics of these models are presented in the following applications.

5.2 Application 1: Managerial Efficiency in Major League Baseball

In Major League Baseball managers are given a set of players by the team they work for. It is the job of each manager to take those players and produce as many wins as possible. Team executives and fans are often concerned with whether or not their manager is producing as many wins as possible. One might be tempted to simply compare the level of production of one manager to all other managers in order to determine how well they are performing. However, comparisons between Major League Baseball managers are difficult to make due to the fact that each manager is given a different set of players. Therefore, if one manager wins more games than another it may be due to either superior managerial abilities or superior players. In order to truly capture a manager's contribution to their team one must take into account the resources that manager has to work with. In other words, evaluations of major league managers should consider not only the output of the manager but also the inputs the manager is given. In his analysis of the role of race in the survival of major league managers Volz (2009) evaluates managerial performance by applying data envelopment analysis to data on major league managers. The following analysis is an updated and expanded presentation of that methodology.

The goal of this analysis is to determine which managers are doing the best job of producing wins given their resources. Therefore, an efficient manager is defined as a manager who produces the maximum possible wins given his inputs. This analysis assumes that the production of wins is constrained by some underlying technology or production function. Once this production function is identified an efficient team is defined as one whose input output combination lies on the production possibilities frontier. The degree of inefficiency can, therefore, be measured as the distance from the observed level of output to the maximum level of output holding the observed inputs constant. In order to compare inefficiencies of managers with different inputs the inefficiency is measured as the percentage of possible output which is produced. This measure is referred to as the output oriented technical efficiency.

In order to calculate this measure of efficiency we must have a production possibilities frontier to compare our observed input output bundles to. The method of data envelopment analysis allows us to construct this frontier based on several reasonable assumptions about the nature of the underlying technology. These assumptions are most easily illustrated graphically for a single input single output case. Therefore, Fig. 5.1 represents the DEA production possibilities frontier for a model in which managers turn one input, player talent, into one output, wins. The first assumption of this model is that all convex combinations of feasible input output bundles are also feasible. In the single input single output case this implies that all points which lie on a line between any two observed points are feasible since those observed points must be feasible. These convex combinations are shown in Fig. 5.1 by the lines connecting each observed manager's input output bundle. The second assumption is that there is free disposability of inputs. This assumption implies that if a level of input can produce a given output then any level of input which is greater than or equal to that level of input in all dimensions can also produce that output. For the single input single output case presented in Fig. 5.1, this implies that all points to the right of a feasible point are also feasible. The third assumption is that of free disposability of output. This assumption implies that if a given level of input can produce some level of output it can also produce a level of output which is less than or equal to that level of output in all dimensions. For the single input single output case presented in Fig. 5.1, this implies that any point which lies below a feasible point is also feasible.

These three assumptions are very reasonable when evaluating Major League managers. The convexity assumption implies that if a manager is given a talent level between that of two other teams they should be capable of producing a number of wins between that of those other teams. The assumption of free disposability of inputs implies that if another team is producing a given level of wins and you have better players you should be able to produce at least that many wins. Lastly, the

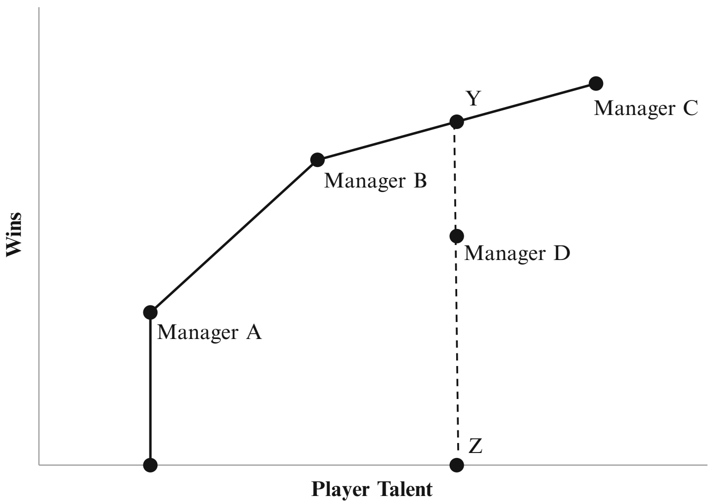


Fig. 5.1 DEA production possibilities frontier

assumption of free disposability of output implies that if a manager can produce a certain level of wins with a given set of players then it should also be possible to produce fewer wins with those same players.

Based on these assumptions, the observed input output bundles can be used to construct a production possibilities frontier like that seen in Fig. 5.1. Once that production possibilities frontier is constructed the percentage of possible wins which are being produced can be calculated. For the one input one output case presented in Fig. 5.1, this is calculated by taking the wins of the manager under consideration and dividing that output by the level of wins directly above that manager on the frontier. For example, the efficiency of Manager D can be calculated as follows. This manager is clearly below the frontier and therefore not producing efficiently. Using a convex combination of managers B and C we can construct an input output bundle represented by point Y. This input output bundle has the same player talent but the maximum wins possible given our assumptions about the production technology. In order to find the efficiency of Manager D we can simply divide the output of Manager D by the hypothetical level of output at point Y. Graphically, this is the distance from D to Z divided by the distance from Y to Z. This represents the percentage of possible wins which Manager D is producing given his level of player talent. This output oriented technical efficiency can then be used to compare amongst managers with different levels of input.

Major League managers are given several different players and therefore will have more than one input into their production of wins. For the multiple inputs case this measure cannot be shown graphically but can still be calculated mathematically. This is accomplished by solving the following maximization problem based on the work of Banker et al. (1984).

$$
\begin{aligned}
\text{Maximize} \quad & \theta \\
\text{Subject to:} \quad & \sum \lambda_i Y_i \geq \theta Y_0 \\
& \sum \lambda_i X_{ij} \leq X_{0j} \\
& \lambda_i \geq 0 \\
& \sum \lambda_i = 1
\end{aligned}
$$

Where i indicates the firm and j indicates the input. Firm 0 is the firm being evaluated.

In this problem θ is a factor by which the observed firms output is multiplied. Maximizing θ implies that we are looking for the largest possible increase in output. The first constraint states that the convex combination of observed firms must produce a level of output which is greater than or equal to that of the firm being evaluated. The second constraint implies that the convex combination of observed firms must have inputs which are less than or equal to that of the firm being evaluated. The constraint that each lambda be greater than or equal to zero eliminates the possibility of any firm having a negative weight in the convex combination of firms. The final constraint that the lambdas must sum to one eliminates the possibility of a scaled up or down version of a single firm's input output bundle. This essentially creates a variable returns to scale production possibilities frontier. Variable returns to scale are appropriate for this application as doubling a manager's inputs will not necessarily double his team's wins.

The resulting θ from this maximization problem will represent the factor by which a firm can increase their output while keeping their inputs at or below their observed level. The inverse of θ is the percentage of possible output which a firm is producing given its level of inputs. This output oriented technical efficiency can be used to compare performances amongst firms with different inputs. If a firm is producing on the production possibilities frontier this measure is equal to one and the firm is considered efficient. For any inefficient firm this measure will be less than one and greater than zero.

For the case of Major League Baseball managers, the output of interest is wins. However, due to the fact that not all managers manage an entire season and not all seasons have had the same number of games wins may not be comparable from one observation to the next. Therefore, the chosen measure of output for this analysis is winning percentage. For this analysis playoff wins are not included. This is due to the fact that managers given a very small level of talent cannot be expected to make the playoffs and therefore will only concern themselves with regular season wins.

A manager's maximum winning percentage will depend on the talent he is given to work with. Therefore, the input of interest is the talent level of the players a manager is given. This talent can come in many forms, such as batting, pitching, fielding, and base running. Baseball analysts have traditionally measured these talents through statistics such as batting average, earned run average, errors, and stolen bases. These statistics measure a player's performance during a specific time period under the guidance of a specific manager. It is likely that these performance measures will be influenced by that manager. It is expected that a good manager will have a positive impact on a player's performance while a bad manager will have a negative impact on a player's performance. If player talent is measured in terms of runs scored then the measure will always overstate the level of talent given to good managers and understate the level of talent given to bad managers.

Ideally, the talent input should be a measure of player potential which is not influenced by the manager being evaluated. It is reasonable to assume that teams pay players based on how they expect them to perform. This level of pay is determined before the player actually plays and therefore should be less influenced by the manager than traditional performance measures. Based on this logic player salaries are used as inputs to the production of winning percentage. Therefore, the output oriented technical efficiency presented here evaluates how efficiently a manager produces wins given his set of player salaries.

Due to data limitations and for simplicity of analysis the salaries are divided into offensive salaries and defensive salaries. For offense, the salaries of the players who played the most games at each infield position and the top three outfielders in games played are summed for each team of each year from 1985 to 2011. American League games are played using a designated hitter. Therefore, they have one more offensive player in the American League requiring the analysis to be

Table 5.1 Baseball player price index

Year	Index	Year	Index
1985	5.66	1999	2.04
1986	8.27	2000	1.69
1987	7.43	2001	1.50
1988	6.88	2002	1.40
1989	6.74	2003	1.38
1990	5.75	2004	1.39
1991	3.78	2005	1.35
1992	3.23	2006	1.22
1993	2.98	2007	1.20
1994	2.93	2008	1.13
1995	2.84	2009	1.11
1996	2.85	2010	1.08
1997	2.42	2011	1.11
1998	2.27	2012	1.00

conducted separately for the National and American Leagues. Most Major League Baseball teams use five starting pitchers which they rotate from game to game. Additionally, teams use a number of relief pitchers throughout the season. Therefore, the defensive input is equal to the salaries of the top five pitchers in terms of games started and the top six pitchers in terms of relief appearances. Player salaries from 1985 to 2012 are publicly available from several online sources including espn.com, usatoday.com, and the Sean Lahman Baseball Database.

Since 1985, there has been a rapid increase in the level of Major League Baseball player salaries. This makes comparisons between inputs from different seasons impossible. Therefore, player salaries must be adjusted for the overall rise in prices. To accomplish this, a price index is created by calculating the average player salary for each season from 1985 to 2012. The 2012 average salary is then divided by that of the other seasons in order to construct the Baseball Player Price Index. This index, presented in Table 5.1, is used to convert all player salaries to 2012 baseball dollars.

A manager's winning percentage is not only affected by the talent he is given but also by the talent on the teams he plays against. The level of talent of opposing teams can be considered a negative input to the production of wins. If two managers have identical teams but play against different competition those managers cannot be expected to win the same number of games. Therefore, a negative input which captures the talent level of the competition is included in the analysis. Major League Baseball teams play the majority of their games against teams within their own divisions. Therefore, a negative input equal to the average total salary of the other teams in a manager's division is also included in the model.

With the inputs and outputs defined the production possibilities frontier can be constructed. This is done by solving the following maximization problem and taking the inverse of the resulting θ value. This model was originally presented in Volz's (2009) analysis of discrimination in Major League Baseball and is based on the methodology developed by Banker et al. (1984).

$$
\begin{aligned}
&\text{Maximize} && \theta \\
&\text{Subject to :} && (1)\ \sum \lambda_i W_i \geq \theta W_0 \\
& && (2)\ \sum \lambda_i O_i \leq O_0 \\
& && (3)\ \sum \lambda_i D_i \leq D_0 \\
& && (4)\ \sum \lambda_i C_i \geq C_0 \\
& && (5)\ \sum \lambda_i = 1 \\
& && (6)\ \lambda_i \geq 0
\end{aligned}
$$

Constraint 1 implies that the combination of other observed win percentages must be greater than the observed winning percentage of the manager being evaluated. Constraints 2 and 3 imply that the combination of offensive and defensive inputs must be less than the inputs of the manager under consideration. Constraint 4 states that the combination of negative competition inputs must be at least as great as the competition faced by the manager being evaluated. Constraint 5 implies variable returns to scale by eliminating scaled up or down versions of one input output bundle. Constraint 6 assures that there are no negative lambdas.

The fact that a manager's own output, W_0, is a feasible level of output assures that the maximum value of θ will be greater than or equal to 1. θ can be interpreted as the multiple by which winning percentage can be increased using a convex combination of observed managers. Therefore, the output oriented technical efficiency, or the percentage of potential wins which are being produced, is the inverse of θ. Because θ is greater than or equal to 1, the technical efficiency will always lie in the closed interval from 0 to 1. A value of 1 will indicate that a manager is producing the maximum winning percentage given his players.

The output oriented technical efficiency is calculated for each manager for each season from 1986 to 2011 for both the American and National Leagues. In order to increase the number of comparison managers, each manager is evaluated against all managers in their league over three seasons. Those seasons include the season being evaluated along with the previous and following seasons. These efficiencies are calculated for each manager in the American and National Leagues with at least 25 consecutive games managed for the years 1986–2011. A ranking of managers from most to least efficient is presented in Table 5.2. The reported technical efficiencies are game weighted averages from the years 1986 through 2011 for all managers with at least 500 games managed within the sample.

These efficiencies allow for an evaluation of any manager's performance during a specific season or over the course of their career. For example, in order to analyze Charlie Manuel's 2008 World Series championship season his performance is evaluated relative to all National League managers from the years 2007, 2008, and 2009. The inverse of the resulting θ from the maximization problem with

Table 5.2 Technical efficiencies 1986–2011 (Minimum of 500 games in sample)

Rank	Manager	Average TE	Games in sample	Rank	Manager	Average TE	Games in sample
1	Joe Girardi	0.962	810	37	Lou Piniella	0.871	3548
2	Jack McKeon	0.955	1446	38	Terry Collins	0.870	1040
3	Ken Macha	0.949	972	39	Jim Lefebvre	0.869	859
4	Bobby Cox	0.948	3270	40	Joe Morgan	0.869	563
5	Joe Maddon	0.944	1001	41	Buddy Bell	0.868	1221
6	Ned Yost	0.935	1248	42	Jim Tracy	0.867	1574
7	Bob Geren	0.933	710	43	Cito Gaston	0.866	1722
8	Fredi Gonzalez	0.928	717	44	Tom Trebelhorn	0.864	923
9	Art Howe	0.928	2266	45	Ozzie Guillen	0.863	1295
10	Bob Brenly	0.927	565	46	Eric Wedge	0.862	1296
11	Willie Randolph	0.924	555	47	Phil Garner	0.859	2035
12	Felipe Alou	0.919	2054	48	Jim Fregosi	0.859	1637
13	Larry Dierker	0.913	783	49	Hal McRae	0.858	872
14	John Gibbons	0.911	610	50	Buck Rodgers	0.857	1172
15	Bruce Bochy	0.910	2736	51	Jeff Torborg	0.856	994
16	Jim Leyland	0.909	3175	52	Bobby Valentine	0.856	2060
17	Ron Gardenhire	0.898	1621	53	Don Baylor	0.855	1317
18	Pete Rose	0.898	560	54	Jerry Manuel	0.853	1388
19	Gene Lamont	0.898	1115	55	Bud Black	0.853	811
20	Jimy Williams	0.897	1701	56	Mike Hargrove	0.851	2363
21	Frank Robinson	0.896	1326	57	Bob Melvin	0.850	1100
22	Grady Little	0.895	648	58	Lloyd McClendon	0.849	782
23	Terry Francona	0.894	1944	59	Don Zimmer	0.846	524
24	Mike Scioscia	0.893	1944	60	Whitey Herzog	0.845	729
25	Tony LaRussa	0.893	4125	61	Johnny Oates	0.840	1544

(continued)

Table 5.2 (continued)

Rank	Manager	Average TE	Games in sample	Rank	Manager	Average TE	Games in sample
26	Charlie Manuel	0.892	1544	62	John Wathan	0.839	646
27	Kevin Kennedy	0.891	582	63	Larry Bowa	0.827	853
28	Joe Torre	0.890	3134	64	Sparky Anderson	0.826	1555
29	Ron Washington	0.888	810	65	Tommy Lasorda	0.824	1629
30	Tom Kelly	0.880	2363	66	John McNamara	0.822	675
31	Clint Hurdle	0.879	1321	67	Tony Muser	0.804	725
32	Buck Showalter	0.879	1935	68	Manny Acta	0.803	734
33	Davey Johnson	0.878	1798	69	Rene Lachemann	0.798	506
34	Dusty Baker	0.877	2852	70	Dallas Green	0.785	633
35	Roger Craig	0.877	1134	71	Jerry Narron	0.766	633
36	Bob Boone	0.873	815	72	Jim Riggleman	0.762	1475

Charlie Manuel's 2008 statistics as W_0, O_0, D_0, and C_0 is his technical efficiency for that year. Charlie Manuel's resulting technical efficiency is equal to .911 for the 2008 season. This means that his winning percentage was 91.1 % of the maximum possible given his levels of talent and competition. Charlie Manuel's game weighted average technical efficiency over his 1544 games in the sample is .892. This makes him the 26th most efficient manager amongst the 72 managers who managed at least 500 games from 1986 to 2011. The average technical efficiencies presented in Table 5.2 range from a high of .962 for Joe Girardi to a low of .762 for Jim Riggleman.

When calculating output oriented technical efficiencies multiple managers are located on the production possibilities frontier each period. Therefore, multiple managers will have a technical efficiency equal to one in each period with no way to rank their performances. Andersen and Petersen (1993) propose a modified model of technical efficiency which is capable of ranking multiple efficient performances. This measure is referred to as super-efficiency and is applied as follows.

Comparisons amongst efficient firms can be made by determining how much an efficient firm's output can be reduced while still remaining efficient. This super-efficiency measure is calculated by running the same model as previously presented with the added constraint that the lambda of the manager under consideration be

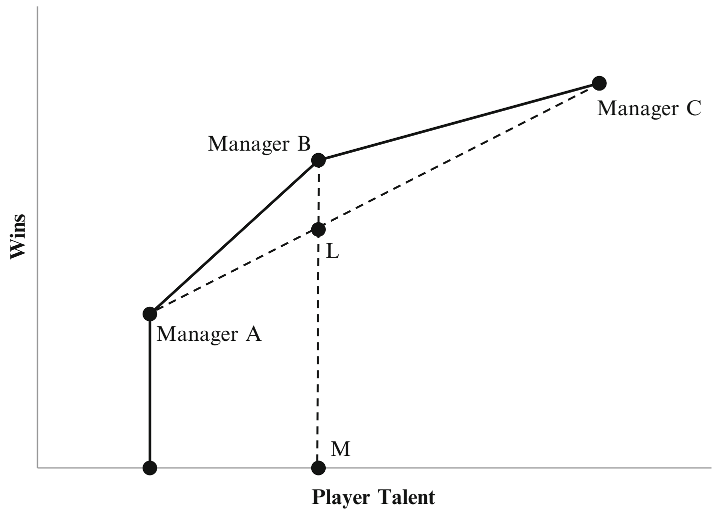


Fig. 5.2 Super-efficiency production possibilities frontier

equal to zero. This results in the production possibilities frontier being based on all observations except that of the manager under consideration.

This model is depicted graphically for the single input single output case in Fig. 5.2. In this example, Manager B is located on the production possibilities frontier and therefore has an output oriented technical efficiency equal to one. In order to calculate the super-efficiency measure for Manager B we compare Manager B to a frontier created using all managers except Manager B. This frontier is shown in Fig. 5.2 by the dashed line from Manager A to Manager C. Manager B's efficiency relative to this frontier is calculated as the distance from Manager B to point M divided by the distance from point L to point M. This value will be greater than one for all firms which lie on the original frontier and less than one for all firms which lie below the original frontier. For inefficient managers, the value of this super-efficiency measure will be the same as that of the output oriented technical efficiency presented previously. The advantage of the super-efficiency measure is that those firms which are efficient will now have a measure greater than one. This allows for comparison between efficient managers and therefore gives a more accurate ranking of managerial performance.

A disadvantage of this methodology is that under certain conditions there may be no feasible solution to the maximization problem. For example, in Fig. 5.2, super-efficiency cannot be calculated for Manager C as there are no observations with more player talent than Manager C. Therefore, we cannot construct a frontier for that level of player talent without using Manager C. The following super-efficiency model is calculated for each manager from 1986 to 2011.

$$\begin{aligned}
&\text{Maximize} \quad \theta \\
&\text{Subject to}: \ (1) \ \sum \lambda_i W_i \geq \theta W_0 \\
&\qquad (2) \ \sum \lambda_i O_i \leq O_0 \\
&\qquad (3) \ \sum \lambda_i D_i \leq D_0 \\
&\qquad (4) \ \sum \lambda_i C_i \geq C_0 \\
&\qquad (5) \ \sum \lambda_i = 1 \\
&\qquad (6) \ \lambda_i \geq 0 \\
&\qquad (7) \ \lambda_0 = 0
\end{aligned}$$

The only difference between this model and the previous model is the addition of constraint 7. Constraint 7 requires that the team under consideration must have lambda equal to zero. As previously stated, this implies that each manager is now compared to a frontier which does not include themselves. This added constraint results in there being no feasible solution for 52 out of the 829 observations. The game weighted average super-efficiencies for each manager from 1986 to 2011 with at least 500 games managed are presented in Table 5.3. Note that when efficient seasons are evaluated based on super-efficiency Joe Girardi is replaced by Jack McKeon as the most efficient manager over the sample period. Jack McKeon's average efficiency increased from .955 to .982. This implies that for some seasons Jack McKeon produced a level of wins which was above the level of wins required to be considered efficient. This Super-Efficiency was not captured in the rankings produced by the previous model. Clearly this information would be useful to any baseball executive or fan looking to compare amongst managers who have been given different levels of talent to work with.

Table 5.3 Super-efficiencies 1986–2011 (Minimum of 500 games in sample)

Rank	Manager	Average SE	Games in sample	Rank	Manager	Average SE	Games in sample
1	Jack McKeon	0.982	1446	36	Joe Morgan	0.871	563
2	Bob Brenly	0.982	565	37	Ozzie Guillen	0.867	1295
3	Joe Maddon	0.981	677	38	Tom Kelly	0.865	2039
4	Ken Macha	0.979	972	39	Clint Hurdle	0.865	1159
5	Joe Girardi	0.961	648	40	Bud Black	0.863	811
6	Bobby Cox	0.956	3270	41	Buck Rodgers	0.863	1172
7	Kevin Kennedy	0.954	582	42	Jim Tracy	0.861	1412
8	John Gibbons	0.941	610	43	Lou Piniella	0.858	3062
9	Felipe Alou	0.934	1586	44	Bobby Valentine	0.856	2060
10	Art Howe	0.930	1456	45	Phil Garner	0.855	1920

(continued)

Table 5.3 (continued)

Rank	Manager	Average SE	Games in sample	Rank	Manager	Average SE	Games in sample
11	Bob Geren	0.930	548	46	Bob Melvin	0.854	1100
12	Larry Dierker	0.926	783	47	Jerry Manuel	0.853	1388
13	Willie Randolph	0.924	555	48	Cito Gaston	0.852	1560
14	Pete Rose	0.911	560	49	Eric Wedge	0.849	1134
15	Ron Gardenhire	0.909	1621	50	Don Zimmer	0.846	524
16	Jimy Williams	0.908	1701	51	Terry Collins	0.846	878
17	Bruce Bochy	0.903	2250	52	Whitey Herzog	0.845	729
18	Tony LaRussa	0.901	4125	53	Tom Trebelhorn	0.843	761
19	Buck Showalter	0.901	1773	54	Bob Boone	0.841	654
20	Terry Francona	0.898	1944	55	Johnny Oates	0.841	1544
21	Joe Torre	0.897	3134	56	John Wathan	0.840	646
22	Davey Johnson	0.897	1798	57	Mike Hargrove	0.840	2039
23	Grady Little	0.895	648	58	Jim Lefebvre	0.839	697
24	Mike Scioscia	0.894	1944	59	Don Baylor	0.835	1155
25	Ned Yost	0.893	763	60	Hal McRae	0.829	724
26	Jim Leyland	0.893	2547	61	Larry Bowa	0.827	853
27	Buddy Bell	0.893	897	62	Sparky Anderson	0.826	1555
28	Charlie Manuel	0.892	1544	63	Tommy Lasorda	0.824	1629
29	Roger Craig	0.889	1134	64	John McNamara	0.822	675
30	Dusty Baker	0.888	2852	65	Lloyd McClendon	0.810	621
31	Ron Washington	0.888	810	66	Dallas Green	0.794	633
32	Jim Fregosi	0.882	1637	67	Jerry Narron	0.766	633
33	Jeff Torborg	0.881	994	68	Jim Riggleman	0.765	1475
34	Gene Lamont	0.881	953	69	Tony Muser	0.747	563
35	Frank Robinson	0.872	1002	70	Manny Acta	0.747	572

5.3 Application 2: Efficient Resource Allocation in Major League Baseball

Unlike a Major League Baseball manager who is given a specific set of players and must make the most of them, Major League teams have the ability to adjust their inputs. Therefore, when it comes to evaluating Major League teams we may want to consider not only how many wins a team could produce but also how they should allocate their resources to accomplish that level of wins. This can be accomplished using a DEA model similar to that presented in the previous evaluation of managers.

As in the previous application, an efficient team is defined as a team which produces the maximum possible output given its level of inputs. One of the major advantages of data envelopment analysis is that multiple outputs can easily be incorporated into the model. When analyzing Major League teams this is particularly useful as there are multiple types of wins a team would like to produce. For the case of a firm producing multiple outputs we can simply define an efficient firm as one which is producing the maximum level of each output possible given its level of inputs. As with the previous application, a frontier can be constructed based on the assumptions of convexity, free disposability of inputs, and free disposability of outputs. A measure of efficiency for multiple outputs can be calculated through the following maximization problem based on the work of Banker et al. (1984).

$$\begin{aligned} \text{Maximize} \quad & \theta \\ \text{Subject to:} \quad & \sum \lambda_i Y_{ij} \geq \theta Y_{0j} \\ & \sum \lambda_i X_{ij} \leq X_{0j} \\ & \lambda_i \geq 0 \\ & \sum \lambda_i = 1 \end{aligned}$$

Where i indicates the firm and j indicates the input or output. Firm 0 is the firm being evaluated.

In this problem θ is a factor by which the observed firms output is multiplied. The goal is to maximize θ which represents an increase in production of all outputs. The difference between this model and the model previously presented is that Y_0 has been replaced with Y_{0j}. This implies that there are multiple outputs and that the value of each output in the convex combination of other firms must exceed that of the firm being evaluated. The second constraint implies that the convex combination of observed firms must have inputs which are less than or equal to that of the firm under consideration. The constraint that each lambda be greater than or equal to one eliminates the possibility of any firm having a negative weight in the linear combination of firms. The final constraint that the lambdas must sum to one eliminates the possibility of a scaled up or down version of a single firms input output bundle. This again creates a variable returns to scale production possibilities frontier. Variable returns to scale are appropriate for this application as doubling a team's inputs will unlikely double the teams output.

The resulting θ from this maximization problem will represent the factor by which a firm can increase all of their outputs while keeping their inputs at or below their observed level. The inverse of this θ is the percentage of possible output which a firm is producing given its level of inputs. This measure is referred to as the output oriented technical efficiency. If a firm is producing efficiently this measure is equal to one. For any firm not producing on the production possibilities frontier this measure will be less than one and greater than zero.

As mentioned earlier, there are multiple types of wins which Major League Baseball teams produce. The first type of win is a regular season win. It is important to produce as many regular season wins as possible as regular season wins determine a team's standing within their division. Therefore, whether a team makes the playoffs and has a chance to win the World Series is based on regular season wins. Additionally, regular season wins increase attendance thereby increasing revenues. Based on this logic regular season winning percentages are used as a measure of output for each team in each year. Winning percentage is chosen over the actual number of wins as it makes comparisons between seasons of slightly different lengths possible.

The second type of win is a playoff win. Some may argue that these wins are even more important than regular season wins as a team can win all of its regular season games but if it does not win any playoff games it cannot win a championship. Therefore, playoff wins for each team in each year are included as a second measure of output. Playoff wins are measured as the combined wins in the League Championship Series and World Series for all teams that made the playoffs each year and zero for the teams who did not make the playoffs. Division Series wins are not included in this analysis as the Division Series did not exist in the early years of the sample. Additionally, the year 1994 is dropped from the analysis as there were no playoffs that year due to a player strike.

Baseball is played with a specific number of players on the field at any given time. Therefore, the number of inputs to the production of wins is the same for each team. However, while each team has, for example, one starting first baseman, the amount of money they allocate towards that first baseman can vary. A team can choose to have a first baseman that costs the league minimum or they can choose to have a first baseman that costs ten million dollars. Therefore, the inputs to production for this analysis are measured as the salaries of the players at each position.

For the positions of first base, second base, third base, shortstop, and catcher the variables are the salary paid to the player on each team who played the most games at that position. Additionally, the American League uses a designated hitter to bat for the pitcher while the National League does not. Therefore, the American League analysis includes an extra input that is the salary of the player who played the most games as the designated hitter.

Due to data limitations some positions are grouped into categories. The outfield variable is the total salary for the three players who played the most games as outfielders. Most teams have five starting pitchers which they rotate each game.

Therefore, the starting pitcher variable is the total salary of the five pitchers who started the most games for each team. Teams have at least six pitchers that are used in relief based on the situation. Therefore, the relief pitcher variable is calculated as the total salary paid to the six pitchers on each team that had the most relief appearances in a given season.

As discussed in the previous application a team's ability to win is not only influenced by the players which it has but also by the talent of the teams it plays against. Therefore, as in the previous application the talent of the competition a team faces is included as a negative input to the production of wins. Since teams play the majority of their games within their own division this competition input is measured as the average total team salary of the other teams in each team's division.

These inputs and outputs were collected for all teams from 1985 to 2012. The salaries are converted to 2012 baseball dollars using the Baseball Player Price Index presented in Table 5.1. With the inputs and outputs defined, technical efficiencies can be calculated for each team based on the following data envelopment analysis model.

$$\begin{aligned} &\text{Maximize} \quad \theta \\ &\text{Subject to}: \ (1)\ \sum \lambda_i \mathrm{WINPER}_i \geq \theta \mathrm{WINPER}_0 \\ &\qquad (2)\ \sum \lambda_i \mathrm{PWIN}_i \geq \theta \mathrm{PWIN}_0 \\ &\qquad (3)\ \sum \lambda_i \mathrm{TOTAL}_i \leq \mathrm{TOTAL}_0 \\ &\qquad (4)\ \sum \lambda_i \mathrm{COMP}_i \geq \mathrm{COMP}_0 \\ &\qquad (5)\ \sum \lambda_i = 1 \\ &\qquad (6)\ \lambda_i \geq 0 \end{aligned}$$

Where 0 indicates the manager being analyzed and i indexes the other managers in the reference group.

In this model, constraints 1 and 2 imply that the convex combination of observations must have at least as high a winning percentage and as many post season wins as the team being evaluated. Constraint 3 implies that the convex combination of observations have total salaries less than or equal to that of the team being evaluated. Constraint 4 implies that the convex combination of observations face a level of competition greater than or equal to that of the team being evaluated. Constraints 5 and 6 ensure variable returns to scale and positive solutions.

Note that while individual inputs were calculated for each position those individual inputs are not included in the constraints of the model. This is because, unlike managers who are given inputs which they must make the best of, teams are free to allocate their budgets however they would like. In other words, teams are constrained by their overall budget but they are free to shift resources between positions. This is why this model includes one constraint which requires that the total team salary of the convex combination be less than that of the team under consideration. It does not have to be the case that each individual position have a lower salary than the team under consideration. This means that the convex combination of teams may have more or less salary at any given position than the team being evaluated so long as the total is not greater. This allows us to compare a team to the relevant convex combination of other teams in order to see if improving their output requires shifting resources from one position to another.

The team under consideration is evaluated in comparison to the teams in their league from that season and the previous and following seasons. This is based on the assumption that teams from adjacent years are operating under comparable production functions while teams separated by several years are not expected to have the same production technology. This results in output oriented technical efficiencies being calculated for all teams in each league from 1986 to 2011. The average results for each franchise from this model are presented in Table 5.4.

Table 5.4 Average technical efficiency (1986–2011)

Rank	Franchise	Average TE
1	New York Yankees	0.909
2	Florida	0.907
3	Boston	0.898
4	Tampa Bay	0.896
5	Atlanta	0.894
6	Oakland	0.892
7	St. Louis	0.883
8	San Francisco	0.883
9	Montreal/Washington	0.881
10	Toronto	0.879
11	Milwaukee	0.871
12	Pittsburgh	0.869
13	San Diego	0.868
14	Houston	0.864
15	Minnesota	0.863
16	Cincinnati	0.860
17	Arizona	0.859
18	Philadelphia	0.855
19	California/Anaheim/Los Angeles	0.854
20	Cleveland	0.853
21	Los Angeles Dodgers	0.850
22	Colorado	0.842
23	New York Mets	0.841
24	Chicago White Sox	0.829
25	Texas	0.809
26	Seattle	0.806
27	Detroit	0.794
28	Chicago Cubs	0.790
29	Baltimore	0.773
30	Kansas City	0.771

The higher the technical efficiency the more efficient a firm is with efficient firms having a technical efficiency equal to one.

As can be seen in Table 5.4, the average output oriented technical efficiencies range from a high of .909 for the New York Yankees and a low of .771 for the Kansas City Royals. This implies that from 1986 to 2011 the New York Yankees produced on average 90.9 % of the regular season and post season wins possible given their player budget. On the opposite end of the spectrum, the Kansas City Royals only produced 77.1 % of the regular season and post season wins possible given their budget.

In addition to evaluating the average efficiency of different franchises over time, these results can be used to evaluate the allocation of resources within a specific team during any particular season. This can be done by comparing the team under consideration to the convex combination of other teams which would produce the maximum output given the team under consideration's budget. While the convex combination of teams must have a lower overall budget they may have more or less money allocated to any individual position. This enables us to identify which positions a team is spending too much or too little money on.

As an example, we can evaluate the 2000 New York Mets season. For the 2000 season, the Mets had an output oriented technical efficiency of .945. This implies that there exists a convex combination of other teams which produced more wins from a lower budget while facing greater competition. Specifically, a convex combination of the 1999 Atlanta Braves and the 2001 Arizona Diamondbacks can be constructed with lower inputs but higher outputs. This convex combination is illustrated in Table 5.5.

The lambdas in Table 5.5 are the weights given to Arizona and Atlanta when constructing the reference team. The last row of Table 5.5 shows the differences between the Mets inputs and outputs and that of the reference team. Note that the convex combination has greater regular season and playoff wins while at the same time using a lower total budget and facing higher competition. This shows that all of the constraints in the DEA model are satisfied. While the reference team must have a lower total salary the model does not required it to have a lower salary at each position. Therefore, by looking at the difference between positions we are able to identify potential strategies the Mets could use to improve their output. For example, the number that stands out in the bottom row of Table 5.5 is that the Mets appear to have paid almost \$19 million more to their catcher than the convex combination of Atlanta and Arizona did. In 2000, the Mets paid over \$20 million (in 2012 baseball dollars) to Mike Piazza, a twelve time all-star and former rookie of the year. On the other hand, the reference team allocates relatively little salary towards catcher but is able to achieve a higher level of wins none the less. This may suggest that the Mets would have been better off reallocating some of their catcher salary towards a position such as starting pitching where they have a lower value than the reference team.

Similar comparisons could be made between any team and the appropriate convex combination of teams. In this manner team executives could use DEA models in order to identify opportunities for improvement through reallocation of

Table 5.5 Detailed analysis of the 2000 mets (Salary values in millions of 2012 baseball dollars)

	λ	Win%	Playoff Wins	Total	Comp	1B	2B	3B	SS	C	OF	SP	RP
2000 NY Mets	–	58.02	5.00	123.5	58.1	7.3	7.2	13.6	5.1	20.5	9.2	39.8	21.0
2001 Arizona	0.32	56.79	8.00	137.4	84.1	4.5	12.1	13.5	6.0	1.9	17.6	40.3	41.5
1999 Atlanta	0.68	63.58	4.00	103.6	51.5	9.7	5.9	8.5	6.1	1.3	13.1	52.6	6.4
Reference team	–	61.39	5.29	114.5	62.0	8.0	7.9	10.1	6.1	1.5	14.5	48.6	17.8
Mets-Reference		−3.36	−0.29	9.0	−3.9	–0.7	–0.7	3.4	–1.0	19.0	–5.3	–8.9	3.2

resources. A limitation of this methodology is that due to labor market constraints teams may not be able to reallocate salaries to the extent they wish. For example, the ability of the Mets to reallocate salary from catcher to pitcher depends on them being able to find and acquire quality pitching talent. It is also possible that they could be stuck with specific players due to long term contracts. However, even if this is the case there is still value in teams knowing what reallocations would be beneficial in case such opportunities were to arise in the future.

References

Andersen P, Petersen NC (1993) A procedure for ranking efficient units in data envelopment analysis. Manage Sci 39(10):1261–1264

Anderson TR, Sharp GP (1997) A new measure of baseball batters using DEA. Ann Oper Res 73:141–155

Banker RD, Charnes A, Cooper WW (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis. Manage Sci 30(9):1078–1092

Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision making units. Eur J Oper Res 2(6):429–444

Chen WC, Johnson AL (2010) The dynamics of performance space of Major League Baseball pitchers 1871–2006. Ann Oper Res 181(1):287–302

Einolf KW (2004) Is winning everything? A data envelopment analysis of Major League Baseball and the National Football League. J Sports Econ 5(2):127–151

Howard LW, Miller JL (1993) Fair pay for fair play: estimating pay equity in professional baseball with data envelopment analysis. Acad Manage J 36(4):882–894

Lewis HF, Sexton TR (2004a) Data envelopment analysis with reverse inputs and outputs. J Product Anal 21(2):113–132

Lewis HF, Sexton TR (2004b) Network DEA: efficiency analysis of organizations with complex internal structure. Comput Oper Res 31(9):1365–1410

Lewis HF, Sexton TR, Lock KA (2007) Player salaries, organizational efficiency, and competitiveness in major league baseball. J Sports Econ 8(3):266–294

Lewis HF, Lock KA, Sexton TR (2009) Organizational capability, efficiency, and effectiveness in Major League Baseball: 1901–2002. Eur J Oper Res 197(2):731–740

Mazur MJ (1994) Evaluating the relative efficiency of baseball players. In: Charnes A et al (eds) Data envelopment analysis: theory, methodology, and applications. Kluwer Academic, Dordrecht, pp 369–391

Miceli TJ, Volz BD (2012) Debating immortality: application of data envelopment analysis to voting for the Baseball Hall of Fame. Manag Decis Econ 33(3):177–188

Sexton TR, Lewis HF (2003) Two-stage DEA: an application to major league baseball. J Product Anal 19(2–3):227–249

Sueyoshi T, Ohnishi K, Kinase Y (1999) A benchmark approach for baseball evaluation. Eur J Oper Res 115(3):429–448

Volz B (2009) Minority status and managerial survival in major league baseball. J Sports Econ 10(5):522–542

Chapter 6
Efficiency and Productivity in the US Property-Liability Insurance Industry: Ownership Structure, Product and Distribution Strategies

J. David Cummins and Xiaoying Xie

Abstract The chapter examines efficiency and productivity of US property-liability (P-L) insurers using data envelopment analysis (DEA). We estimate pure technical, scale, cost, revenue and profit efficiency over the period 1993–2011. Insurers' adjacent year total factor productivity changes and their contributing factors are also investigated. In particular, we explore the relationship of insurers' efficiency with their ownership structure, product and distribution strategies. Regression analyses are also performed to explore the relationships between firm characteristics, efficiency and productivity. The results indicate US P-L insurance industry has improved its efficiency and productivity over time. Insurers' product strategy, distribution system, and diversification strategy are important determinants of insurers' efficiency and productivity, along with other firm characteristics.

Keywords US property-liability insurance industry • Efficiency • Productivity • Data envelopment analysis (DEA) • Malmquist index • Organizational form • Mutuals • Distribution systems

6.1 Introduction

In the past two decades, we have observed significant changes in the economy of the United States, which has also altered the landscape in the US property-liability (P-L) insurance marketplace. The industry has been going through the innovations in computer and communications technologies since the early 1990s that have enabled insurers to adopt more efficient insurance marketing, underwriting and claims adjusting strategies. The prevailing use of social media and telematics in

J.D. Cummins
Temple University, 617 Alter Hall, 1801 Liacouras Walk, Philadelphia, PA 19122, USA
e-mail: cummins@temple.edu

X. Xie (✉)
California State University-Fullerton, 800 N. State College Blvd., Fullerton, CA 92834, USA
e-mail: xxie@fullerton.edu

J. Zhu (ed.), *Data Envelopment Analysis*, International Series in Operations Research & Management Science 238, DOI 10.1007/978-1-4899-7684-0_6

recent years has also brought both opportunities and challenges to insurance companies in their operation and product development. The 1990s and 2000s also witnessed several of the most significant catastrophic losses to property-liability insurers in history. Increased exposure to catastrophic risk, such as mega-earthquakes, more frequent and severe hurricanes and tornados, and unexpected man-made risks have raised new challenges regarding solvency ability and the capacity of the industry. Meanwhile, the development of alternative risk transfer mechanisms and instruments have challenged the use of traditional insurance products to transfer risks, which places additional pressure on the insurance industry to provide quality services to maintain its profit margin.

With the recent financial crisis, the insurance industry has been under increasing public scrutiny for its impact on the health of the financial system (Cummins and Weiss 2014). The drastic changes in the post-crisis legal and regulatory environment undoubtedly will gradually reshape this industry. Regulators have adopted new reinsurance collateral requirements, made adjustments to the model holding company law to pay more attention to enterprise risks, passed the Risk Management and Own Risk and Solvency Assessment (RMORSA) model law to assess enterprise risk management (ERM) and capital levels, and are developing a corporate governance model law. All these changes make US regulations more in line with international standards and comply with the International Monetary Fund's Financial Sector Assessment Program (FSAP) requirements. These changes may help reduce entry barriers for foreign insurers in the US market and bring more competition to existing insurers.

In such a rapidly changing environment, firms have the opportunity to improve their efficiency and productivity by adapting to new technologies, innovating products and services offered, and improving corporate risk management and corporate governance. Meanwhile, they also must face the challenge of rapid adaptation and keeping up with the competition. How do these changes affect firms' efficiency and productivity? The US P-L insurance industry provides us a particularly interesting environment for in-depth analyses. In this chapter, we provide a comprehensive analysis of the performance of US P-L insurers, both static and dynamic. We estimate pure technical, scale, cost, revenue, and profit efficiency for firms in the US P-L insurance industry over the period 1993–2011. Insurers' adjacent year total factor productivity and its contributing factors are also investigated. We explore the relationship of insurers' efficiency with their ownership structure, public status, product, and distribution strategies in detail. Regression analyses are performed to explore the relationships between firm characteristics, efficiency and productivity. This is one of the few papers in the literature that analyzes firm profit efficiency using the most updated data and long time series to track the performance of insurers in the industry.

Since the US P-L insurance industry is a competitive industry,[1] it makes great sense to use benchmarking techniques to measure firm performance and identify

[1] As shown in Cummins and Xie (2013), the Herfindahl index of US P-L insurance industry is below 400 in all years during 1993–2009.

potential improvement by comparing a firm's performance with other firms in the industry. Because data envelopment analysis (DEA), a non-parametric technique (Cooper et al. 2000), has become a widely-used frontier efficiency methodology in measuring firm performance, we have adopted this technique to estimate various efficiencies for insurance firms by constructing "best practice" frontiers for production, cost, and revenue for each year of the sample period.

In this study, we use an input-oriented approach for analyzing cost efficiency and its components, where firms minimize input usage and costs conditional on output levels. Total factor productivity estimation is also input-oriented. Meanwhile, we estimate revenue efficiency by using the output-oriented approach, where firms maximize output quantities and revenues conditional on inputs. Use of the input orientation for inputs and costs and the output orientation for revenues is consistent with the micro-economic theory of the firm.

Analysis of both cost efficiency and revenue efficiency enables us to provide a complete picture of the success of insurers in achieving the economic goals of cost minimization and revenue maximization. In addition, profit efficiency is estimated to capture the overall success of firms in maximizing profits. We decompose cost efficiency into pure technical, scale, and allocative efficiency, which allows us to provide valuable information on the sources of (in)efficiencies in the industry. We also decompose total factor productivity into technical efficiency change and technical change to identify the driving force of productivity improvement.

The remainder of the chapter is organized as follows. Section 6.2 provides an intuitive explanation of the DEA efficiency methodology and mathematical programming for various types of efficiency, along with a literature review on its application in the property-liability insurance industry. Section 6.3 presents the results of the overall efficiency and productivity in the US P-L insurance industry and discusses their implications for insurers' operations. The section also analyzes the relationship between firm efficiency and ownership structure; discusses how firms' efficiency varies with product strategies; presents the impact of marketing strategies on firm efficiency, and conducts a regression analysis to find determinants of efficiencies and productivity for insurers. Section 6.4 summarizes and concludes.

6.2 Frontier Efficiency Methodology and Its Application in the P-L Insurance Industry

This section begins by discussing the efficiency and productivity estimation methodology utilized in this chapter, followed by a summary of literature on the application of DEA efficiency in P-L insurance. We then present and discuss the measurements of inputs, input prices, outputs, and output prices used in our analysis.

6.2.1 *Estimation Methodology of DEA Frontier Efficiency and Malmquist Index*

The essence of efficiency analysis is to separate production units that perform well from those that perform poorly. This is done by estimating "best practice" efficient frontiers consisting of the fully efficient firms in an industry and comparing all firms in the industry to the frontier. Two major categories of methodologies—parametric and non-parametric—have been developed to estimate efficient frontiers. The parametric approaches require the specification of a functional form for the frontier (e.g., a cost or revenue function) and also require assumptions about the probability distributions of the error terms of the model. Non-parametric approaches do not require assumptions about either the functional form of the frontier or the error term distributions. The primary advantage of the parametric approach is that firms are allowed to deviate from the frontier due to random error as well as inefficiency, whereas the non-parametric approaches measure all departures from the frontier as inefficiency. The disadvantage of the parametric approach is that efficiency estimates can be confounded by specification error if the wrong assumptions are made about the functional form and error term distributions.

Although there was disagreement in the early financial institutions' efficiency literature about whether the parametric or non-parametric approach was more appropriate, research by Cummins and Zi (1998) for insurance and Casu et al. (2004) for banking shows that parametric and non-parametric approaches generally produce consistent results. Moreover, Cummins and Zi (1998) show that non-parametric approaches tend to correlate better with conventional performance measures such as return on equity. In addition, non-parametric approaches such as DEA provide a particularly convenient way to decompose overall productivity and efficiency and estimate scale economies of firms. DEA is also quite appealing intuitively because it implements micro-economic theory by constructing efficient frontiers specified as optimization problems whereby decision making units (DMUs) minimize costs and maximize revenues or profits (Cooper et al. 2000).

DEA has excellent asymptotic statistical properties. Banker (1993) shows that DEA is equivalent to maximum likelihood estimation, with the specification of the production frontier in DEA as a non-parametric monotone and concave function instead of a parametric form linear in parameters. DEA estimators are consistent and converge faster than estimators from other frontier methods (Kneip et al. 1998; Grosskopf 1996). DEA estimators are also unbiased if we assume no underlying model or reference technology (Kittelsen 1995). Banker and Natarajan (2008) also prove that DEA can yield consistent estimators for contextual variables in a two-stage regression with DEA efficiency as the dependent variable.

Six types of efficiency are estimated by DEA in this study: technical efficiency (production frontiers with constant returns to scale), pure technical efficiency (production frontiers with variable returns to scale), non-increasing returns to scale (NIRS) technical efficiency (production frontiers with non-increasing returns to scale), cost efficiency, revenue efficiency, and profit efficiency. The cost, revenue, and profit efficiencies are estimated under the assumption of constant returns to scale, i.e., firms are allowed to deviate from the frontiers due to scale inefficiency as well as pure technical and allocative inefficiency.

All six types of efficiency are estimated for each individual firm in each year of the sample period. To standardize notation, we assume that there are N firms in the industry in a given year and that each firm uses a maximum of k inputs $x = (x_1, x_2, \ldots, x_k) \in \Re_+^k$ to produce a maximum of m outputs $y = (y_1, y_2, \ldots, y_m) \in \Re_+^m$. The input price vector is $w = (w_1, w_2, \ldots, w_k) \in \Re_+^k$, and the output price vector is $p = (p_1, p_2, \ldots, p_m) \in \Re_+^m$. DEA is then used to construct a frontier (production, cost, or revenue) for each year such that all observed firms lie on or below the frontier. Firms lying on the frontier are regarded as "best practice" firms, and those below the frontier are inefficient relative to the "best practice" firms in that given year.

6.2.1.1 Production Frontiers and Technical Efficiency

We employ input-oriented distance functions introduced by Shephard (1970) to estimate production frontiers. The input-oriented distance function relies on the *input attainability* assumption that all output vectors can be obtained from the rescaling of any non-zero input vectors. Let the correspondence $y \to V(y) \in \Re_+^k$ denote the production technology that transforms inputs into outputs. Then for any $y \in \Re_+^m$, $V(y)$ is the subset of all input vectors $x \in \Re_+^k$ that yields at least y. The input distance function is therefore defined as

$$D(x, y) = \sup\left\{\phi : \left(\frac{x}{\phi}, y\right) \in V(y)\right\} = \frac{1}{\inf\{\theta : (\theta x, y) \in V(y)\}} \tag{6.1}$$

Farrell's (1957) input-oriented technical efficiency (TE) is then defined as

$$\mathrm{TE}(x, y) = \frac{1}{\mathrm{D}(x, y)} = \inf\{\theta : (\theta x, y) \in V(y)\} \tag{6.2}$$

Technical efficiency reflects a firm's ability to minimize the inputs utilized to produce a given bundle of outputs, i.e., $\mathrm{TE}(x, y)$ represents the radial contraction in inputs for a firm to produce a given output vector y if it operated on the production frontier. Fully efficient firms lie on the production frontier and have efficiency scores equal to 1.0. Inefficient firms have efficiency scores between 0 and 1, and $1-\mathrm{TE}(x, y)$ is the inefficiency due to not adopting the best production technology.

Farrell technical efficiency can be measured with respect to production frontiers characterized by constant returns to scale (CRS), variable returns to scale (VRS), and non-increasing returns to scale (NIRS) (Aly et al. 1990). Three DEA production frontiers are then estimated by solving linear programming problems as specified below.

Input-oriented CRS technical efficiency (TE^I_{CRS}) for firm j is estimated by solving:

$$\begin{aligned} &\mathrm{TE}^{\mathrm{I}}_{\mathrm{CRS}}(x,y) = \mathrm{Min}\ \theta^I_{CRS} \\ \text{Subject to} \quad &\sum_{\mathrm{j=1}}^{\mathrm{N}} \lambda_{\mathrm{j}} y_{ij} \geq y_{ij} \qquad \forall\ \mathrm{i} = 1, \ldots, \mathrm{m} \\ &\sum_{\mathrm{j=1}}^{\mathrm{N}} \lambda_{\mathrm{j}} x_{rj} \leq \theta^I_{CRS} x_{rj} \quad \forall\ \mathrm{r} = 1, \ldots, \mathrm{k} \\ &\lambda_{\mathrm{j}} \geq 0 \qquad \forall\ \mathrm{j} = 1, \ldots, \mathrm{N} \end{aligned} \tag{6.3}$$

Input-oriented VRS technical efficiency (pure technical efficiency) for firm j is estimated using the same problem setup, except that a convexity constraint is imposed by including the following condition in the optimization: $\sum_{\mathrm{j=1}}^{\mathrm{N}} \lambda_{\mathrm{j}} = 1$. Input-oriented NIRS technical efficiency is estimated by changing the constraint to $\sum_{\mathrm{j=1}}^{\mathrm{N}} \lambda_{\mathrm{j}} \leq 1$.

Overall input-oriented technical efficiency, TE^I_{CRS}, can be expressed as the product of pure technical efficiency and scale efficiency. A firm has achieved pure technical efficiency if it operates on the VRS frontier and has achieved full technical efficiency if it operates on the CRS frontier. Scale efficiency for a given firm is measured as the total technical efficiency not explained by pure technical efficiency, i.e., $\mathrm{SE}^I(x,y) = \frac{\mathrm{TE}^I_{\mathrm{CRS}}(x,y)}{\mathrm{TE}^I_{\mathrm{VRS}}(x,y)}$, where SE^I = input-oriented scale efficiency and TE^I_{VRS} = input-oriented VRS technical efficiency. If $SE^I = 1$, the firm is on the CRS frontier; and if $SE^I < 1$, the firm is not on the CRS frontier. If $SE^I < 1$ and $TE^I_{VRS} \neq$ NIRS efficiency, the firm operates with increasing returns to scale (IRS); and if $SE^I < 1$ and TE^I_{VRS} = NIRS efficiency, the firm is characterized by decreasing returns to scale (DRS) (Aly et al. 1990).

6.2.1.2 Cost Frontiers and Cost Efficiency

The cost efficiency program utilized in this chapter is input-oriented. Here, the objective is to minimize cost by choosing input quantities while holding constant the input prices w and output quantities y. The linear programming problem for firm j is:

$$
\begin{aligned}
&\mathrm{C}(x,y) = \mathrm{Min}_{\lambda,x_j} \sum_{\mathrm{r}=1}^{\mathrm{k}} w_{rj} x_{rj} \\
&\text{Subject to} \quad \sum_{\mathrm{j}=1}^{\mathrm{N}} \lambda_{\mathrm{j}} y_{ij} \geq y_{ij} \qquad \forall\, \mathrm{i} = 1, \ldots, \mathrm{m} \\
&\qquad\qquad\quad \sum_{\mathrm{j}=1}^{\mathrm{N}} \lambda_{\mathrm{j}} x_{rj} \leq x_{rj} \qquad \forall\, \mathrm{r} = 1, \ldots, \mathrm{k} \\
&\qquad\qquad\qquad\qquad \lambda_{\mathrm{j}} \geq 0 \qquad \forall\, \mathrm{j} = 1, \ldots, \mathrm{N}
\end{aligned}
\tag{6.4}
$$

The solution $x_j^* = \left\{x_{1j}^*, x_{2j}^*, \ldots, x_{rj}^*, \ldots, x_{kj}^*\right\}$ is the cost-minimizing vector of input quantities for firm *j*. The cost efficiency (CE) of the firm is the ratio of optimal cost over actual cost, i.e., $\mathrm{CE}(x,y) = \frac{\sum_{r=1}^{k} w_{rj}\, x_{rj}^*}{\sum_{r=1}^{k} w_{rj}\, x_{rj}}$, where $x_j = \left\{x_{1j}, x_{2j}, \ldots, x_{rj}, \ldots, x_{kj}\right\}$ is the observed input quantity vector of the firm. The value of CE is bounded between 0 and 1, with CE = 1 for fully cost efficient firms.

Cost efficiency is the product of technical efficiency and allocative efficiency, where allocative efficiency measures the success of the firm in choosing cost minimizing combinations of inputs and is calculated as: $\mathrm{AE}^I(x,y) = \frac{\mathrm{CE}(x,y)}{\mathrm{TE}^I_{\mathrm{CRS}}(x,y)} = \frac{\mathrm{CE}(x,y)}{\mathrm{TE}^I_{VRS}(x,y)*\mathrm{SE}^I(x,y)}$. Even if a firm produces on the production frontier, it is not fully cost efficient if it is not allocatively efficient. The complement 1-AE measures the cost inefficiency of a firm due to its failure to adopt the optimal combinations of inputs, given the input prices and output quantities.

6.2.1.3 Production Frontiers and Revenue Efficiency

Another important objective of firms is to maximize revenues by choosing optimal output quantities while holding constant output prices *p* and input quantities *x*. Revenue efficiency is output-oriented and based on Shephard's (1970) output distance function. The output distance function is based on the assumption of "output attainability" that *all* input vectors are feasible in the production of any rescaled nonzero output vector. The linear programming problem to estimate the revenue efficiency for firm *j* is specified as:

$$
\begin{aligned}
&\mathrm{R}(x,y) = \mathrm{Max}_{\lambda, y_j} \sum_{\mathrm{i=1}}^{\mathrm{m}} p_{ij} y_{ij} \\
&\text{Subject to} \quad \sum_{\mathrm{j=1}}^{\mathrm{N}} \lambda_j y_{ij} \geq y_{ij} \quad \forall\, \mathrm{i} = 1, \ldots, \mathrm{m} \\
&\qquad\qquad\quad \sum_{\mathrm{j=1}}^{\mathrm{N}} \lambda_j x_{rj} \leq x_{rj} \quad \forall\, \mathrm{r} = 1, \ldots, \mathrm{k} \\
&\qquad\qquad\quad \lambda_j \geq 0 \qquad \forall\, \mathrm{j} = 1, \ldots, \mathrm{N}
\end{aligned}
\tag{6.5}
$$

The solution $y_j^* = \left\{ y_{1j}^*, y_{2j}^*, \ldots, y_{ij}^*, \ldots, y_{mj}^* \right\}$ is the revenue-maximizing vector of output quantities for firm j. The revenue efficiency of the firm is then given as the ratio of observed revenue over optimal revenue: $\mathrm{RE}(x,y) = \frac{\sum_{i=1}^{m} p_{ij} y_{ij}}{\sum_{i=1}^{m} p_{ij} y_{ij}^*}$, where $y_j = \left\{ y_{1j}, y_{2j}, \ldots, y_{ij}, \ldots, y_{mj} \right\}$ is the firm's observed output quantity vector. Efficient firms have $\mathrm{RE} = 1$, and inefficient firms have RE between 0 and 1.

6.2.1.4 Profit Efficiency

Since the ultimate goal of a firm is to maximize profits, we estimate profit efficiency for our sample firms to show the net effects of cost and revenue efficiency. The profit efficiency metric we use in this chapter is defined in terms of the firm's actual profits and optimal profits, i.e., the profits that could be obtained if the firm were fully efficient. The profit efficiency measure differs from cost and revenue efficiency by allowing the firm to optimize over both inputs and outputs. The DEA profit efficiency model we utilize is from Cooper et al. (2000), based on a model originally proposed in Färe et al. (1985). The model solves:

$$
\begin{aligned}
&\underset{x_j, y_j}{Max} \quad p_j^T y_j - w_j^T x_j \\
&\text{subject to:} \quad Y\lambda_j \geq y_j \\
&\qquad\qquad\quad X\lambda_j \leq x_j \\
&\qquad\qquad\quad \lambda_j \geq 0
\end{aligned}
\tag{6.6}
$$

where the i^{th} row of y_j and r^{th} row of x_j in the objective are defined by:

$$y_{ij} = \sum_{q=1}^{N} y_{iqj}\lambda_{qj}, \ \ i = 1, \ldots, m$$

and

$$x_{rj} = \sum_{q=1}^{N} x_{rqj}\lambda_{qj}, \ \ r = 1, \ldots, k \ , \tag{6.7}$$

On the right hand side of (6.7), y_{iqj} = the qj^{th} element of the i^{th} row of Y (output matrix of all firms in the industry), and $x_{rqj} = qj^{th}$ element of the r^{th} row of X (input matrix of all firms in the industry). As in Cooper et al. (2000), we estimate profit *inefficiency* as:

$$\pi_j = \left(p_j^T y_j^* - w_j^T x_j^*\right) - \left(p_j^T y_j - w_j^T x_j\right) \tag{6.8}$$

where y_j^* and x_j^* are, respectively, the m element optimal output vector and k element optimal input vector obtained by solving the problem in expression (6.6). Thus, (6.8) provides a measure of the "profits lost (in the form of an 'opportunity cost') by not operating in a fully efficient manner" (Cooper et al. 2000, p. 222).

To scale profit efficiency to be more consistent with our other efficiency measures, we express profit inefficiency as a ratio, where we normalize π_j by dividing by the sum of actual costs and revenues, $\left(p_j^T y_j + w_j^T x_j\right)$ (see Cooper et al. 2004). We do not use optimal or actual profits as the denominator because optimal profits can be 0 and actual profits can be 0 or negative. Therefore, unlike the *efficiency* ratios, profit *inefficiency* does not have to be between 0 and 1. The most efficient firms will have profit *inefficiency* equal to 0, while profit inefficient firms will have an efficiency score greater than 0.

6.2.1.5 Total Factor Productivity and Its Decomposition

The Malmquist index approach has been widely used to analyze total factor productivity (TFP) changes of firms over time. One advantage of Malmquist index is that that it permits the separation of technical change from efficiency change (Färe et al. 1994) (hereafter FGNZ 1994) and is also DEA based. There are several ways of decomposing the Malmquist index into its components based on different assumptions of returns to sale of the production frontier (see FGNZ 1994; Ray and Desli 1997; Simar and Wilson 1998).[2] In this chapter we adopted FGNZ's (1994)

[2] For detailed discussion, see Cummins and Xie (2008).

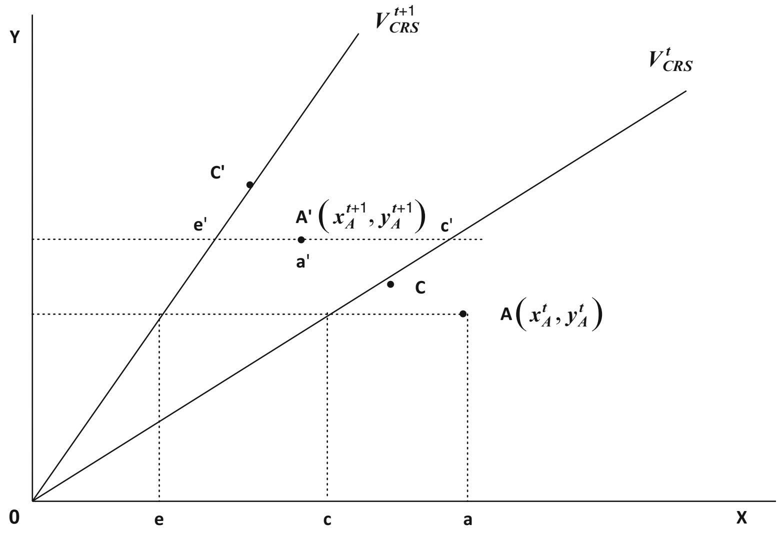


Fig. 6.1 Malmquist productivity measurement, single input-single output firm

constant returns to scale (CRS) technology assumption and decompose the TFP change of a firm into its two primary components: technical change (the shift in the production frontier over time), and efficiency change (the shift in the firm's efficiency relative to the production frontier over time).

We utilize input-oriented Malmquist productivity indices in this chapter. The methodology is demonstrated in Fig. 6.1, which shows production frontiers for a single-input (X), single-output (Y) industry.

The CRS production frontier in period t is represented by the line 0 V_{CRS}^{t}, and the CRS production frontier in period $t+1$ is represented by the line 0 V_{CRS}^{t+1}. Firm A produces at point A in period t and produces at point A' in period $t+1$. During the two periods, this firm has experienced two changes: (1) the firm is using better technology in period $t+1$. As shown in the figure, this firm's input–output combination in period $t+1$ would have been infeasible using period t technology; (2) The firm is operating closer to the production frontier in period $t+1$ than in period t, indicating a technical efficiency gain between the two periods.

Input-oriented distance functions are defined as

$$D^t\left(x_j^s, y_j^s\right) = sup\left\{\phi_j^s : \left[\frac{x_j^s}{\phi_j^s}, y_j^s\right] \in V^t\left(y_j^s\right)\right\} = \frac{1}{inf\left\{\theta_j^s : \left(\theta_j^s x_j^s, y_j^s\right) \in V^t\left(y_j^s\right)\right\}} \tag{6.9}$$

where $D^t(x_j^s, y_j^s)$ is the input-oriented distance function for firm j in period s relative to the production frontier in period t with CRS technology. We can then measure the productivity changes of firm j over time by allowing the $s \neq t$. For example, in

Fig. 6.1, D^t and D^{t+1} represent the distance function relative to the CRS production frontier at time t and $t+1$, respectively, where (x_j^t, y_j^t) is input–output combination of firm j at time t, and $\left(x_j^{t+1}, y_j^{t+1}\right)$ is its input–output combination at time $t+1$.

Then, from Fig. 6.1, $D^t\left(x_A^t, y_A^t\right) = \frac{0a}{0c}$, $D^{t+1}\left(x_A^{t+1}, y_A^{t+1}\right) = \frac{0a'}{0e'}$. Likewise, $D^t\left(x_A^{t+1}, y_A^{t+1}\right) = \frac{0a'}{0c'}$, $D^{t+1}\left(x_A^t, y_A^t\right) = \frac{0a}{0e}$. $D^t(x_A^t, y_A^t)$ and $D^{t+1}\left(x_A^{t+1}, y_A^{t+1}\right)$ compare the period t ($t+1$) input–output vector to the same period's production frontier and must have value ≥ 1. However, if the frontiers shift over time, the distance function $D^t\left(x_A^{t+1}, y_A^{t+1}\right)$ and $D^{t+1}\left(x_A^t, y_A^t\right)$ can be <1, implying that a given period's input–output combination is infeasible using the other period's technology.

Based on FGNZ (1994), a Malmquist index can be defined relative to either the technology in period t (written as M^t) or the technology in period $t+1$ (written as M^{t+1}),

$$M^t = \frac{D^t\left(x_A^t, y_A^t\right)}{D^t\left(x_A^{t+1}, y_A^{t+1}\right)}, \quad \text{or} \quad M^{t+1} = \frac{D^{t+1}\left(x_A^t, y_A^t\right)}{D^{t+1}\left(x_A^{t+1}, y_A^{t+1}\right)}, \tag{6.10}$$

where M^t measures productivity growth between periods t and $t+1$ using the period t technology as benchmark, while M^{t+1} measures productivity growth between periods t and $t+1$ using the period $t+1$ technology as a benchmark. To avoid an arbitrary choice of reference technology, the Malmquist total factor productivity index is then defined as the geometric mean of M^t and M^{t+1},

$$M\left(x_A^{t+1}, y_A^{t+1},\ x_A^t, y_A^t\right) = \left[M^t * M^{t+1}\right]^{\frac{1}{2}} = [(0\text{a}/0\text{c})(\ 0\text{c}'/0\text{a}')(\ 0\text{a}/0\text{e})(\ 0\text{e}'/0\text{a}')]^{\frac{1}{2}} \tag{6.11}$$

FGNZ (1994) decompose the above Malmquist index into technical efficiency change (EFFCH) and technical change (TECHCH) as follows:

$$\text{EFFCH} = \frac{D^t\left(x_A^t, y_A^t\right)}{D^{t+1}\left(x_A^{t+1}, y_A^{t+1}\right)} \tag{6.12}$$

$$\text{TECHCH} = \left[\left(\frac{D^{t+1}\left(x_A^{t+1}, y_A^{t+1}\right)}{D^t\left(x_A^{t+1}, y_A^{t+1}\right)}\right)\left(\frac{D^{t+1}\left(x_A^t, y_A^t\right)}{D^t\left(x_A^t, y_A^t\right)}\right)\right]^{\frac{1}{2}} \tag{6.13}$$

A Malmquist index >1 (<1) implies total factor productivity growth (decline), and similar interpretation for technical efficiency change and technical change.

6.2.2 Summary of Literature on Application of DEA Efficiency in Insurance

The DEA efficiency methodology has seen growing acceptance in insurance research. Cummins and Weiss (2013) provide the most comprehensive and up-to-date literature review on the application of frontier efficiency and productivity methods to the insurance industry. The paper summarizes the concept of efficiency and productivity, including both DEA and econometric frontier efficiency models; the issues of measurement of input, output, and their prices in insurance efficiency estimation; and studies using frontier efficiency methodologies in the insurance industry. In this section, we provide a short literature review on the application of DEA efficiency in the *property-liability* (P-L) insurance area.

6.2.2.1 Economies of Scale and Scope

The DEA method provides a convenient way to measure economies of scale and economies of scope of firms, two essential aspects in industrial organizations. A representative work on economies of scope, by Cummins et al. (2010), studies the economies of scope of the US insurance industry (both life-health and property-liability industries) by looking at both cost scope economies and revenue scope economies to examine the comparative advantage of conglomeration vs. strategic focus.

DEA is also used in studies of economies of scale. Some papers have estimated scale efficiency as a by-product when studying other important issues. For example, Cummins and Nini (2002) estimate scale efficiency in their study of capital utilization in the P-L industry using data from 1993 to 1998. Cummins and Xie (2008) estimate scale economies for US P-L insurers for the period 1994–2003 when studying mergers and acquisitions for the industry. The most recent study is by Cummins and Xie (2013). Using data from 1993 to 2009, the paper provides a detailed analysis of economies of scale of the US P-L insurance industry and analyzes firm characteristics that are associated with their returns to scale. Studies that use DEA method to estimate scale economies for P-L insurers in countries other than the US include Fukuyama and Weber (2001) (Japan), Diacon et al. (2002) (Europe), Mahlberg and Url (2003) (Austria), Jeng and Lai (2005) (Japan), Cummins and Rubio-Misas (2006) (Spain), and Mahlberg and Url (2010) (Germany).

6.2.2.2 Firm Organizational Form, Corporate Governance, Distribution Systems

DEA efficiency and productivity has been increasingly considered a superior performance measure to traditional financial ratios, and we have observed more studies recently using DEA to analyze the performance of firms *per se*, to identify the reasons for inefficiency (input/output slacks), or to examine the relationship between firm characteristics and efficiency. The commonly investigated characteristics include firms' organizational form, distribution strategy, and corporate governance. For example, Cummins et al. (1999b) investigate the expense preference and managerial discretion hypotheses of stock and mutual insurers in the US P-L insurance market using the cross-frontier DEA method. Similarly, Cummins et al. (2004) test the expense preference and efficient structure hypotheses for Spanish life and non-life insurers. Additional studies that test the relationship between organizational form and firm efficiency using DEA methods include Brockett et al. (2004, 2005) (U.S.), Barros et al. (2005) (Portugal), and Jeng and Lai (2005) (Japan).

Changes in insurers' ownership form have provided a convenient way to test the relationship between firm performance and ownership form. The commonly observed ownership changes include mutualization and demutualization, issuing initial public offerings (IPOs), or delisting from the stock exchange. A few recent studies have explored these issues using DEA frontier efficiency measures. For example, Xie (2010) examines changes in public listing status (insurers issuing IPOs) during 1994–2005 and firm cost and revenue efficiency improvement for US P-L insurers. Chen et al. (2011) study the impact of demutualization on the cost efficiency and total factor productivity change of US P-L insurers and find a positive effect.[3]

A couple of recent studies have used DEA method to investigate the relationship between firm performance and corporate governance. For example, Wang et al. (2007) examine the impact of corporate governance structure on the efficiency performance of insurance companies in Taiwan's P-L and life insurance markets. He et al. (2011) study the impact of CEO turnover on firm efficiency and productivity in the US P-L insurance industry. Huang et al. (2011) examine the relationship between efficiency of US P-L insurers and firm corporate governance structure, especially the impact of the passage of Sarbanes-Oxley Act in 2002. Leverty and Grace (2012) investigate the effects of managers on firm efficiency when firms are in financial distress.

There are also studies applying the DEA method to investigate the effectiveness of distribution channels and agent remuneration methods. For example, Ma et al. (2013a, b) study the relationship between the usage of contingent commission and firm efficiency and the impact of the abandonment of such practices on firm efficiency and productivity change.

[3] More studies are on life-insurer demutualization such as Jeng et al. (2007), Erhemjamts and Leverty (2010), and Xie et al. (2011).

6.2.2.3 DEA Efficiency and Regulatory Change of Insurance Market

Insurance markets in many countries and regions have experienced deregulation in the 1990s and early 2000s, which has attracted a set of studies investigating the relationship between deregulation and firm efficiency in those markets. Many papers have adopted the DEA method in measuring firm efficiency, as DEA works better than econometric methods for small samples. Representative work in this category includes Mahlberg and Url (2003) (Austria), Mahlberg and Url (2010) (Germany), and Cummins and Rubio-Misas (2006) (Spain), among many others.

6.2.2.4 DEA Efficiency and Mergers and Acquisitions

Though not yet common, the DEA efficiency method has begun to influence the literature on firm performance after mergers and acquisitions (M&As). The two recent papers include Cummins and Xie (2008) and Cummins and Xie (2009), which use frontier efficiency in measuring firm performance and performance changes of US P-L insurers that have engaged in M&A activities. Davutyan and Klumpes (2009) study consolidation and efficiency in the major European insurance markets using a non-discretionary inputs approach and find an overall beneficial effect on the efficiency of target firms in the industry.

6.2.2.5 DEA Efficiency and Cross-Country Study of Firm Performance

With the globalization of the world economy and the openness of financial markets in many countries, more studies have emerged to compare cross-country performance of insurance industries. The DEA method has been adopted in these studies as well. Eling and Luhnen (2010) compare efficiency of both life and non-life insurers in 36 non-US countries and also compare the results using different efficiency measurement methods. Biener and Eling (2011) study the efficiency performance of micro insurance plans of emerging countries. Biener and Eling (2012) examine the relationships of organization form and efficiency for both life and non-life insurance industries in 21 countries from northern America and the European Union, and Huang and Eling (2013) conduct an efficiency comparison of the non-life insurance industry in the BRIC (Brazil, Russia, India and China) countries.

6.2.2.6 DEA Efficiency and Risk Management

Recent literature has applied DEA efficiency analysis to gauge the effects of firm risk management strategies. Grace et al. (2010) conduct a survey of risk management practices in the insurance industry and examine the impact of enterprise risk management on firm efficiency. They find that firms with chief risk officers (CROs), dedicated risk committees, and risk management entities are more cost efficient than firms without such arrangements.

6.2.3 *Inputs and Outputs Used in the Current Study*

6.2.3.1 Outputs and Output Prices

Cummins and Weiss (2013) summarize the three methods of measuring outputs in insurance frontier efficiency study—the asset (intermediation) approach, the user-cost approach, and the value-added approach.[4] In this study, we adopt a modified version of the value-added approach to define insurance outputs (Berger and Humphrey 1992; Cummins and Weiss 2000).

We define outputs based on the three principal types of services provided by property-liability insurers: *risk pooling and risk bearing, real financial services, and financial intermediation services*. The actuarial, underwriting, claims settlement, and other expenses incurred in operating the risk pools are major components of value-added relating to risk pooling and risk bearing. Real financial services include risk surveys, coverage program design, recommendations regarding deductibles and policy limits, and loss prevention and loss reduction services. The value-added of the intermediation function of P-L insurers is represented by the net interest margin between the rate of return earned on assets and the rate credited to policyholders.

Since detailed transaction data on insurers is not publicly available, the quantity of P-L insurance output is proxied by the present value of real losses incurred (Berger et al. 1997; Cummins and Weiss 2000).[5] Losses incurred are the total amount of losses expected to be pooled and redistributed by the insurers as a result of their providing insurance coverage and is a good proxy for the amount of risk pooling conducted. It is also a good proxy for the amount of real services provided, since these services are highly correlated with aggregate losses. In P-L insurance, because lines of coverage offered by insurers have different risks and payout schedules, we group together lines with similar characteristics.

---

[4] The intermediation approach discussed in Cummins and Weiss (2013a) is the standard intermediation approach applied in many earlier insurance and banking studies. A few authors have (e.g., Brockett et al. 2005) have adopted arbitrary output definitions which they have erroneously called the “intermediation approach.” Leverty and Grace (2010) compare the value-added approach and the Brockett et al. (2005) so-called financial intermediation approach for measuring output in P-L insurer efficiency studies and find that the measurements from these two approaches are not always consistent, with the value-added approach being more closely related to traditional measures of firm performance such as profitability and insolvency risk.

[5] As shown in Cummins and Weiss (2013a), the use of the present value of real losses incurred to measure insurance output has become the dominant approach in the recent P-L insurance efficiency literature.

Four insurance outputs are calculated: personal lines short-tail losses, personal lines long-tail losses, commercial lines short-tail losses, and commercial lines long-tail losses. The tail refers to the length of the loss redistribution period, as defined by Schedule P of the National Association of Insurance Commissioners (NAIC) regulatory statements. The payout proportion of a loss is calculated from data in Schedule P of *Best's Aggregates and Averages* using the chain-ladder method (Lemaire 1985, 1995). Another method frequently used to estimate the payout tail is the Taylor separation method (Taylor 2000). Loss discounting factors are computed from US Treasury yield curves released by the Federal Reserve Board of Governors.[6] The quantity of the intermediation output is measured by the average of the beginning and end-of-year invested assets. The values of losses incurred and invested assets are deflated to real 2000 values using the consumer price index (CPI).

In insurance economics, the value-added of insurance outputs is measured by the Pratt-Arrow concept of the insurance premium (risk premium) (Arrow 1971). In practice, this value-added is the loading of the insurance premium. Accordingly, the price of insurance output is defined as premiums per $1 of present value of incurred losses and loss adjustment expenses.[7]

A few authors have argued that using losses incurred as an output may distort efficiency scores because unexpected increases in losses due to unforeseen catastrophes or adverse shocks may increase the output amount when inputs stay the same, which may result in higher efficiency scores in such cases (Brockett et al. 2004, 2005). A defense to this argument is that using losses as output is justified by the economic theory of insurance; that is, insurers bear residual risks, and policyholders are willing to pay more than expected losses to transfer losses to insurers. It is also justified by the actuarial/statistical theory of insurance, i.e., the objective of insurance is to provide a mechanism for loss pooling and diversification, and a valuable function of insurance is to pay losses even when losses are larger than expected due to catastrophes. Contrary to Brockett et al.'s contention, insurance output really is higher when insurers pay catastrophe claims.

Even though paying unexpected claims is an important role of insurers, losses are still subject to an "errors-in-variables" problem because realized losses have a random component. For example, if losses are larger than expected, then insurers are measured as providing more output. This is not necessarily problematic because insurers have in fact provided more services if losses are higher than expected. Nevertheless, the randomness of losses remains a potential concern, although it is arguably less serious for a non-parametric methodology than for an econometric methodology. Accordingly, we conduct the efficiency estimation using adjusted

[6] We utilize the constant maturity Treasury yields obtained from the Federal Reserve Economic Data (FRED) database maintained by the Federal Reserve Bank of St. Louis. Yields are obtained by linear interpolation for maturities where constant maturity yields are not published in FRED.

[7] Let P_i denotes the price of insurance output i, PE_i denotes the real premiums earned of insurance output i, and LLE_i denotes the real present value of losses and loss adjustment expenses incurred of insurance output i, then $P_i = (PE_i - LLE_i)/LLE_i$.

(smoothed) incurred losses and prices. The smoothing procedure is designed to adjust for errors in variables while still giving insurers credit for paying unexpected claims. A smoothing procedure was also adopted for insurance output prices because we noticed that some insurers had rather extreme values of the unadjusted prices.[8]

The price of the intermediation output is defined as the expected return on invested assets. Invested assets are divided into two categories—stocks and interest-bearing assets (mainly bonds and short-term debt instruments). The price of the intermediation output is calculated as the weighted average expected investment return equal to the expected return on stocks weighted by the proportion of invested assets in stocks plus the expected return on interest-bearing assets weighted by the proportion of the portfolio in this asset type. The expected return on stock assets is calculated as the average 30-day Treasury bill rate in year t plus the long-term (1926 to the end of the preceding year) average market risk premium on large company stocks from Ibbotson Associates (2012). This approach assumes that insurers hold equity portfolios with a market beta coefficient of 1.0. The expected return for interest-bearing assets is estimated as their realized income return in year t, because their expected return is generally close to the actual income return. The realized return on interest-bearing assets equals the total net investment income of the insurer, minus dividends on stocks, divided by the average amount of interest-bearing assets during the year. Thus, the price of the intermediation output differs across insurers.

6.2.3.2 Inputs and Input Prices

We define insurance inputs in four categories—administrative labor, agent labor, materials and business services (including physical capital), and financial equity capital. Because detailed information on number of employees or hours worked is not available by company, we impute the quantities of administrative labor, agent labor, and materials and business services from the dollar value of related expenses. That is, the quantity of an input is defined as the current dollar expenditures related to this input divided by its *current* price. The price of this input is calculated as its *current* price deflated by the CPI, with 2000 as the base year. Thus, the product of the input quantity and the input price equals the constant dollar expenditure on the input.

Current dollar expenditures for administrative labor input are defined as the sum of salaries, payroll taxes, and employee benefits in an insurer's regulatory

[8] See Cummins and Xie (2008) for details of the smoothing procedure. Leverty and Grace (2010) propose an expected value of loss (E(L)) approach in measuring output and compare it with the present value of loss (PV(L)) approach. Expected value of output is estimated by taking the average loss ratio over the previous 3 years and multiplying it with current premiums earned for specific line(s) of business. They find out that the E(L) approach yields slightly lower mean efficiency than the PV(L) approach.

statements. Current dollar expenditures for agent labor input are the sum of net commissions, brokerage fees, and allowances to agents. Current dollar expenditures for materials and business services are calculated as the difference between total expenses incurred and the total administrative and agent labor expenses of the insurer.

The price of administrative labor comes from the US Department of Labor average weekly wage rate for P-L insurance companies (Standard Industrial Classification (SIC) 6331). The category became NAICS 524126 (North American Industry Classification System) in 2001. The price of agent labor comes from the US Department of Labor average weekly wage rate for insurance agents (SIC 6411, NAICS 524210 since 2001). National average weekly wage rates are used here to reduce missing observations. [9] All of these wage variables are deflated to real 2000 values by the CPI to obtain the real prices of the inputs. The current price of the materials and business services input is calculated as a weighted average of price indices for business services from the component indices representing the various categories of expenditures from the expense page of *Best's Aggregates and Averages*. The base year of the price index is 2000. Price indices are obtained from the US Department of Labor and Bureau of Economic Analysis (BEA).

Financial equity capital is considered as an important input, consistent with modern theories of the firm and financial institutions efficiency research (e.g., McAllister and McManus 1993; Berger et al. 1997; Hughes and Mester 1998; and Cummins and Nini 2002). In the financial theory of insurance pricing, insurance is viewed as a risky debt; and the financial equity of insurance companies plays an important role in reducing insolvency risk. We define the financial equity capital of an insurer by summing up the statutory policyholders' surplus and reserves required by statutory accounting principles (SAP) but not recognized by generally accepted accounting principles (GAAP). The quantity of this input is measured by the real value of the average of the beginning and end-of year capital, deflated by the 2000 CPI.

It is ideal to use the market return of equity capital as its price. However, because the majority of insurers are not publicly traded, market equity returns are not observed for most firms in the database. Several approaches measuring the cost of capital are discussed in Cummins and Weiss (2000, 2013). In this chapter, paralleling Cummins et al. (1999a), we estimate the cost of capital for traded insurers in various A.M. Best financial rating categories using the Fama-French three-factor model (Cummins and Phillips 2005) and assign costs of capital to non-traded insurers based on their A.M. Best ratings.[10]

[9] Alternative definitions of input prices include using the home state wage rate for administrative labor and the premium weighted average of state weekly wage rates for agent labor (e.g., Cummins and Nini 2002). Prior literature has shown that using the alternative labor price variables does not materially affect the efficiency results (Cummins et al. 1999).

[10] Robustness checks carried out in various papers (Cummins et al. 1999; Cummins and Nini 2002; Cummins and Xie 2008) show that the results of the analysis are generally not affected by alternative cost of capital assumptions.

6.3 Efficiency and Productivity in the US P-L Industry

6.3.1 Data and Sample Selection

The primary data used in our study are drawn from regulatory annual statements filed by insurers with the NAIC over the period 1993–2011. The decision-making units used in the study consist of groups of affiliated insurers under common ownership and unaffiliated single insurers. Originally, the sample consisted of all groups and unaffiliated insurers for which data are available from the NAIC. We then eliminate firms with zero or negative net worth, premiums, or inputs. Since firms that are extremely small are atypical and may bias the estimation, we eliminate firms whose assets are below $1.5 million. Lastly, we only retain firms with stock, mutual, or reciprocal organizational forms in the sample and eliminate others, such as risk retention groups, US Lloyds, and state workers' compensation fund programs. The final sample used to estimate efficiency consists of 14,592 firms over the entire sample period—an average of 768 firms per year.

6.3.2 Average Efficiency and Variation Across Firms

The results of the DEA efficiency estimation are presented in Table 6.1. Despite fluctuation from year to year, Table 6.1 shows a slight efficiency improvement in this industry over time. For example, the cost efficiency level rose from 45 % in year 1993 to 50 % in year 2011, with the best year, at 56 %, taking place in 1999. Scale efficiency also increased slightly from year 1993 to year 2011, as does allocative efficiency. Pure technical efficiency hardly improved during the sample period. Revenue efficiency improves slightly, from 40 % in 1993 to 45 % in 2011. Profit *inefficiency* of the industry drops significantly from 76 % in 1993 to 56 % in 2011.

The average cost efficiency of the industry is 51 % over the sample period. This implies that P-L insurers, on average, could have reduced costs by 49 % by operating on the production frontier and choosing their input bundles correctly. Decomposing cost efficiency into pure technical, scale, and allocative efficiency provides further information on the sources of cost inefficiency. The average pure technical efficiency of the industry is 74 % over the sample period. A possible reason for the technical inefficiency might be that many insurers failed to adapt to the rapidly changing technology during the sample period, such that they lagged behind their frontier peers. The average allocative efficiency of the industry is 78 %, suggesting that firms could reduce their costs by 22 % if they had used the optimal input combinations.

The average scale efficiency of the industry is 90 %, indicating that 10 % of the inputs were wasted by the industry because the firms did not produce at the optimal scale. Scale efficiency of the industry did not improve very much over the sample

Table 6.1 Average efficiency in the US property-liability insurance industry (1993–2011)

Year	N	Cost	Pure technical	Scale	Allocative	Technical	Revenue	Profit inefficiency
1993	805	0.45	0.75	0.85	0.73	0.64	0.40	0.76
		(0.16)	(0.20)	(0.16)	(0.15)	(0.21)	(0.21)	(0.61)
1994	860	0.47	0.75	0.87	0.74	0.65	0.45	0.61
		(0.16)	(0.19)	(0.14)	(0.13)	(0.20)	(0.21)	(0.46)
1995	864	0.49	0.75	0.88	0.75	0.65	0.45	0.59
		(0.18)	(0.20)	(0.14)	(0.17)	(0.20)	(0.20)	(0.43)
1996	850	0.49	0.76	0.90	0.74	0.68	0.48	0.56
		(0.17)	(0.19)	(0.12)	(0.16)	(0.19)	(0.21)	(0.39)
1997	826	0.44	0.74	0.88	0.70	0.65	0.38	0.99
		(0.17)	(0.20)	(0.14)	(0.16)	(0.20)	(0.23)	(0.80)
1998	807	0.54	0.75	0.91	0.80	0.68	0.45	0.57
		(0.18)	(0.20)	(0.13)	(0.15)	(0.20)	(0.22)	(0.45)
1999	771	0.56	0.77	0.91	0.81	0.69	0.46	0.51
		(0.18)	(0.18)	(0.12)	(0.14)	(0.19)	(0.20)	(0.38)
2000	730	0.55	0.77	0.91	0.79	0.70	0.47	0.53
		(0.17)	(0.18)	(0.11)	(0.14)	(0.19)	(0.21)	(0.40)
2001	700	0.54	0.76	0.90	0.80	0.68	0.42	0.65
		(0.19)	(0.19)	(0.12)	(0.15)	(0.20)	(0.22)	(0.51)
2002	717	0.48	0.72	0.89	0.77	0.64	0.42	0.74
		(0.17)	(0.21)	(0.13)	(0.16)	(0.20)	(0.21)	(0.54)
2003	706	0.51	0.74	0.89	0.78	0.66	0.44	0.65
		(0.18)	(0.20)	(0.12)	(0.16)	(0.20)	(0.22)	(0.54)
2004	726	0.53	0.74	0.91	0.80	0.67	0.44	0.68
		(0.17)	(0.19)	(0.12)	(0.16)	(0.19)	(0.22)	(0.59)
2005	731	0.53	0.75	0.91	0.79	0.68	0.48	0.54
		(0.17)	(0.19)	(0.11)	(0.16)	(0.18)	(0.21)	(0.41)
2006	758	0.49	0.71	0.92	0.76	0.65	0.43	0.67
		(0.16)	(0.20)	(0.11)	(0.14)	(0.19)	(0.20)	(0.49)
2007	778	0.50	0.71	0.91	0.79	0.64	0.42	0.70
		(0.17)	(0.20)	(0.12)	(0.16)	(0.20)	(0.21)	(0.53)
2008	771	0.49	0.71	0.90	0.78	0.63	0.39	0.74
		(0.17)	(0.20)	(0.13)	(0.14)	(0.19)	(0.21)	(0.51)
2009	742	0.53	0.72	0.91	0.83	0.65	0.48	0.49
		(0.16)	(0.20)	(0.10)	(0.15)	(0.19)	(0.19)	(0.29)
2010	745	0.55	0.73	0.92	0.84	0.67	0.51	0.40
		(0.17)	(0.19)	(0.10)	(0.14)	(0.19)	(0.19)	(0.23)
2011	705	0.50	0.73	0.89	0.79	0.64	0.45	0.56
		(0.15)	(0.20)	(0.12)	(0.15)	(0.18)	(0.19)	(0.33)
All years	768	0.51	0.74	0.90	0.78	0.66	0.44	0.63
		(0.17)	(0.20)	(0.13)	(0.16)	(0.19)	(0.21)	(0.50)

Note: This table presents the average efficiency scores of the U.S. property-liability insurance industry during the period 1993–2011. Standard deviation is presented in the parentheses. The sample includes Mutual and Stock firms only. Seven types of efficiencies: pure technical, scale, allocative, technical, cost, and revenue efficiency as well as profit inefficiency, are estimated using data envelopment analysis (DEA)

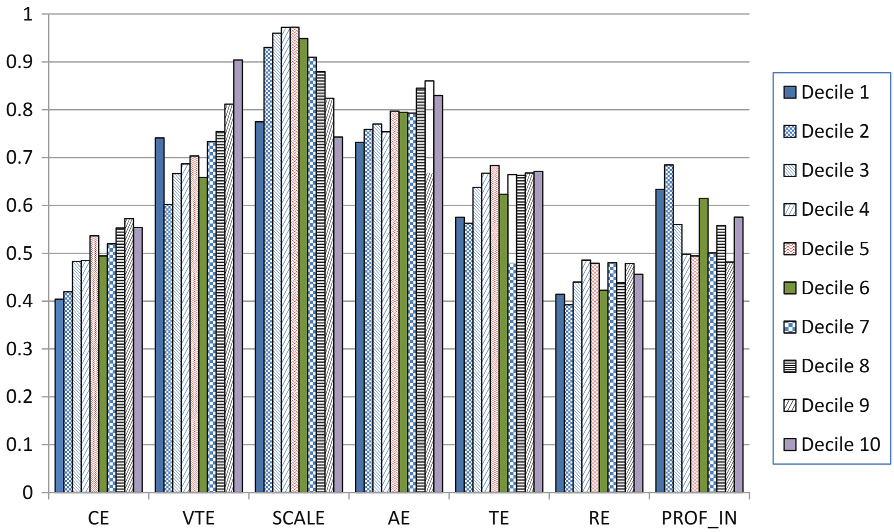


Fig. 6.2 Efficiency by asset size decile, year 2011

period, despite the radical technology innovations the economy has experienced. The average scale efficiency was 85 % in 1993 and 89 % in 2011. One possibility is that technical progress caused the optimal operating scale to rise to a new level. It is also possible that, while some insurers were successful in achieving optimal scale by incorporating new technologies into their production, the performance of others worsened by failing to turn technology into productivity.

The average revenue efficiency of the industry is 44 % during the sample period. This implies that P-L insurers, on average, have the potential to increase revenue by 56 % by operating on the production frontier and choosing their output bundles correctly. The average profit inefficiency of the industry during the sample period is 0.63. Interestingly, we observe an improvement in profit efficiency of the industry right after the financial crisis. The average profit *inefficiency* scores were the lowest in years 2009 and 2010. During these two years, the industry has restructured by freezing hiring and laying off existing employees and incorporating new technologies such as social media into its operations, which probably have reduced the cost and increased revenue efficiency of the industry, thus leading to higher profitability.

The efficiencies are graphed by size decile in Fig. 6.2, where size is based on total assets and decile 1 is the smallest size decile.[11] Cost efficiency is shown furthest to the left in the figure, followed by the decomposition into pure technical, scale, and allocative efficiency. Revenue efficiency and profit *inefficiency* are then shown. Figure 6.2 shows that cost efficiency is monotonically increasing in firm

[11] The figure illustrates the relationship between efficiency and asset size for the year 2011. Using other years' data reaches exactly the same conclusion.

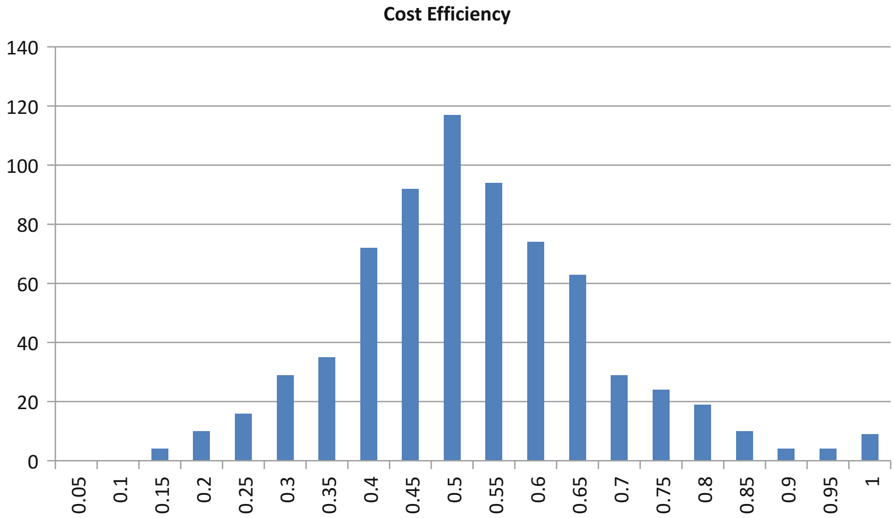


Fig. 6.3 Cost efficiency: US P-L insurers, frequency by efficiency score, 2011

size. This finding is consistent with previous insurance efficiency studies (Gardner and Grace 1993; Cummins and Weiss 1993; Cummins and Zi 1998; Cummins 1999). Scale efficiency peaks in the median deciles and is lower for relatively small and relatively large firms. Large firms on average lose efficiency due to sub-optimal scale, but they are compensated by having higher pure technical and allocative efficiency, leading to the monotonic relationship between size and cost efficiency.

There is no clear relationship between size and revenue efficiency, except that the firms in the two smallest size deciles are the least revenue efficient. Revenue efficiency peaks in deciles 4 and 5 and shows no clear pattern thereafter. Regarding profit inefficiency, overall, there appears to be a negative relationship between firm size and profit inefficiency. Firms in the two smallest size deciles are the least profit efficient, but the clear negative relationship disappears for other deciles. Firms in asset deciles 4, 5, 7, and 9 appear to be the most profit efficient.

As efficiency scores in general increased over the sample period, we provide more detailed analyses of efficiency based on 2011 results. Figures 6.3, 6.4, 6.5, 6.6, and 6.7 demonstrate the frequency of firms by efficiency score ranges for various types of efficiency. The distribution of cost efficiency is shown in Fig. 6.3. Only a few firms (94 firms) have cost efficiency scores lower than 0.35. A majority of firms' cost efficiency scores are between 0.40 and 0.65. About 99 firms have cost efficiency scores greater than 0.70. The overall distribution is bell-shaped. The distribution of allocative efficiency is shown in Fig. 6.4, which is more left skewed. It is noteworthy that there are some firms (71 firms) with extremely low allocative efficiency scores (smaller than 0.60). The average allocative efficiency score is 0.79, most firms fall within the range of 0.85–0.95, and 73 firms have achieved full allocative efficiency. As allocative efficiency measures a firm's ability in choosing

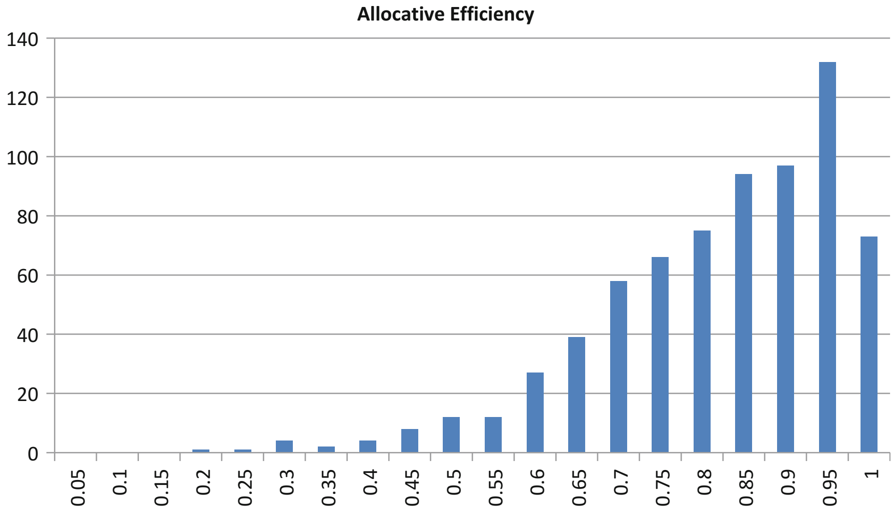


Fig. 6.4 Allocative efficiency: US P-L insurers, frequency by efficiency score, 2011

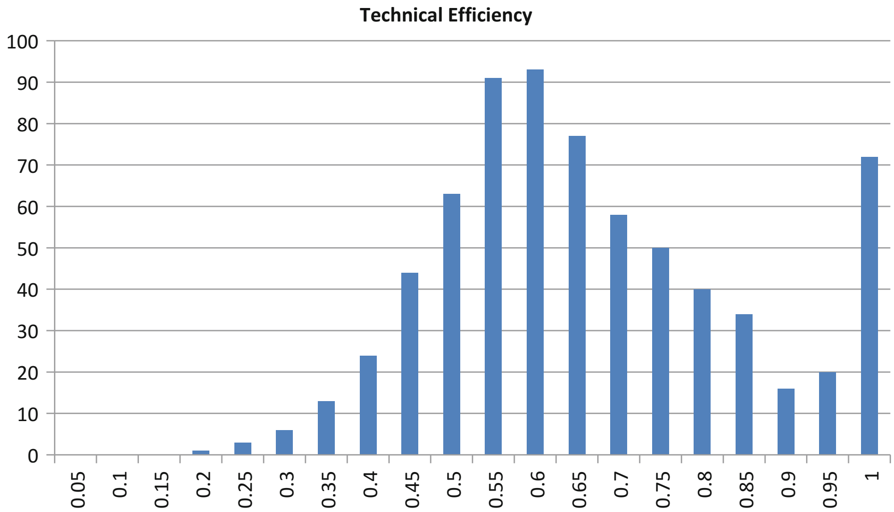


Fig. 6.5 Technical efficiency: US P-L insurers, frequency by efficiency score, 2011

the cost-minimizing combination of inputs, the results show that the majority of firms in the US P-L insurance industry are generally successful in minimizing costs.

Despite the success in input allocation, many firms still fail to achieve cost efficiency because of their low technical efficiency scores. The distribution of technical efficiency is shown in Fig. 6.5. About 6.7 % of firms have technical

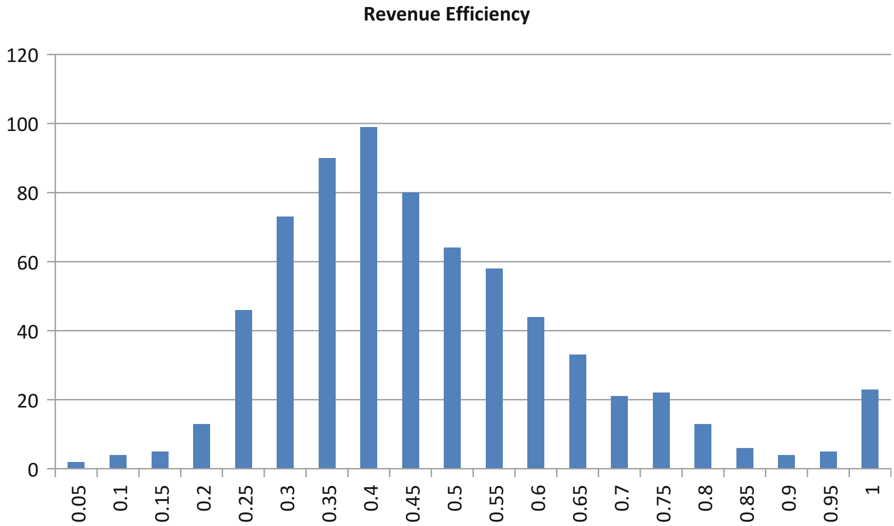


Fig. 6.6 Revenue efficiency: US P-L insurers, frequency by efficiency score, 2011

efficiency scores lower than 0.40, and about 22 % of firms have technical efficiency scores lower than 0.5. A majority of firms' technical efficiency scores fall into the range of 0.55–0.65, with 72 firms being able to achieve full technical efficiency. The rapid change in computer and communication technology during the sample period actually widens the technical efficiency gap between those firms who adapted early and well and those who did not. To many insurers, rapid technological change poses challenges in underwriting, marketing, and claims adjusting services. Slow adaptation may reduce the productivity of a firm and widen the gap between the firm and firms on the industry frontier.

The distribution of revenue efficiency is shown in Fig. 6.6, which is slightly right skewed. The results show that US P-L industry needs to endeavor more to raise its overall revenue efficiency. A majority of firms in the industry have a revenue efficiency score between 0.3 and 0.5, with only 32 firms (5 % of firms) having revenue efficiency scores greater than 0.9. This demonstrates that most firms should be able to choose a more optimal combination of business lines in maximizing their revenues, given their existing input structure.

Figure 6.7 shows the distribution of profit inefficiency. The lower the profit inefficiency score, the more profit efficient the firm is. 55 firms (7.8 % of firms) in the 2011 sample have a profit inefficiency score of zero, demonstrating their ability to maximize profits conditional on their existing outputs and inputs combinations. Other firms have the potential to improve their profits by optimizing over all m-outputs and/or k-inputs.

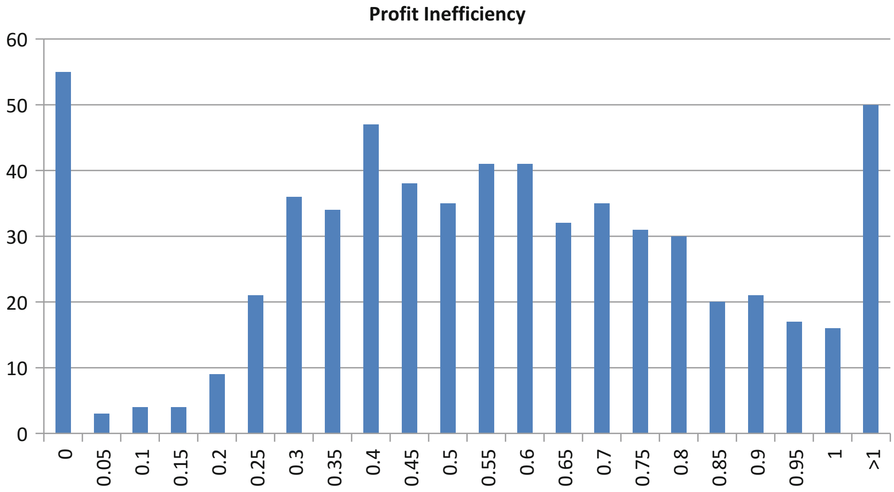


Fig. 6.7 Profit inefficiency: US P-L insurers, frequency by efficiency score, 2011

6.3.3 Productivity of the US P-L Industry and Variation Across Firms

Table 6.2 presents the adjacent-year Malmquist productivity changes of the US P-L insurance industry. We also show the productivity change from 1993 to 2011, comparing the production frontiers in the beginning and ending years of the sample period. Using the adjacent-year comparisons allows us to maintain the largest sample in the analyses, since firms are only required to present in two consecutive years of data in order to be included in this analysis.

The results show that the industry has experienced significant productivity improvement over time. The Malmquist total factor productivity index for the 1993–2011 comparison is 1.173, indicating a 17.3 % improvement in productivity over the sample period. The index is also significantly greater than 1 for 12 of the 18 adjacent-year comparisons and is never significantly less than 1, suggesting that productivity either has improved or stayed the same in each of the adjacent-year comparisons.

Based on the 1993–2011 Malmquist indices, the gains in total factor productivity were primarily attributable to technical change (1.165), i.e., the insurance production frontier shifted in a favorable direction over the sample period such that insurers were able to produce more output for any given input level in 2011 than in 1993. However, technical efficiency change (1.015) was neutral over the sample period, suggesting that technical efficiency neither improved significantly nor deteriorated for firms surviving over the 1993–2011 period.

The conclusions based on 1993–2011 Malmquist indices may suffer a problem of survivor bias. A close look at the adjacent year comparison is able to provide more information. The adjacent-year comparisons also show that technical change

Table 6.2 Constant returns to scale (CRS) Malmquist productivity indices, 1993–2011

Years	N	TFP change	Technical efficiency change	Technical change
1993–1994	725	1.02**	1.05***	0.98**
		(0.22)	(0.19)	(0.19)
1994–1995	764	1.02***	1.04***	0.99**
		(0.18)	(0.17)	(0.13)
1995–1996	753	1.03	1.06***	0.97
		(0.58)	(0.20)	(0.48)
1996–1997	742	1.01**	0.97***	1.05***
		(0.19)	(0.19)	(0.14)
1997–1998	719	1	1.07***	0.94***
		(0.27)	(0.21)	(0.18)
1998–1999	689	1	1.05***	0.96***
		(0.20)	(0.19)	(0.17)
1999–2000	647	1.03***	1.02***	1.01**
		(0.20)	(0.16)	(0.10)
2000–2001	612	1.04***	0.99**	1.06***
		(0.22)	(0.14)	(0.19)
2001–2002	626	1.06***	0.95***	1.12***
		(0.32)	(0.18)	(0.24)
2002–2003	617	1.05***	1.07***	0.99
		(0.17)	(0.22)	(0.13)
2003–2004	641	1.04***	1.05***	1
		(0.21)	(0.20)	(0.14)
2004–2005	641	1.01	1.05***	0.96***
		(0.15)	(0.17)	(0.10)
2005–2006	687	1	0.97***	1.05***
		(0.20)	(0.20)	(0.14)
2006–2007	702	1.03***	1	1.03***
		(0.23)	(0.19)	(0.09)
2007–2008	697	1.02**	1.01	1.02***
		(0.26)	(0.24)	(0.13)
2008–2009	682	1	1.06***	0.95***
		(0.18)	(0.20)	(0.12)
2009–2010	680	1.05***	1.04***	1.01***
		(0.18)	(0.18)	(0.08)
2010–2011	660	1.11***	0.98**	1.14***
		(0.28)	(0.21)	(0.20)
1993–2011[a]	295	1.173***	1.015	1.165***
		(0.42)	(0.33)	(0.25)

Note: This table presents the Malmquist productivity indices for the U.S. property-liability insurance industry. The sample includes Mutual and Stock firms only. The Malmquist indices are based on the Fare et al. (1994) Constant returns to scale (CRS) decomposition of total factor productivity (TFP) change. We estimate adjacent year TFP changes for firms in the sample. t-tests are used to test whether the mean is significantly different from 1. Sign test and Signed rank test return similar significance level. **Significant at the 5 % level; ***significant at the 1 % level
[a]Malmquist analysis, 1993 vs. 2011

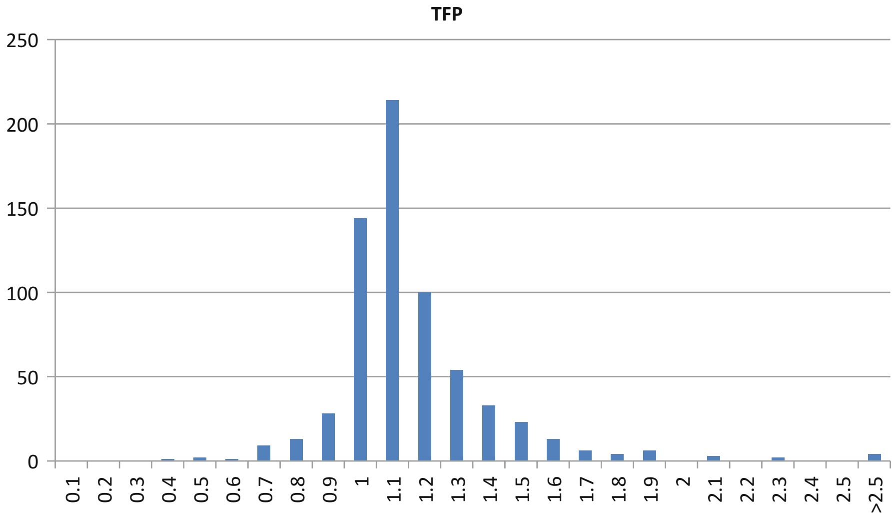


Fig. 6.8 Total factor productivity: US P-L insurers, frequency by score, 2010–2011

was important contributors to total factor productivity gains for insurers. For technical change, the Malmquist index is significantly greater than 1 in 9 of 18 comparisons and significantly less than 1 in 6 comparisons, four of which occur before the year 2000. The technical efficiency change index is significantly greater than 1 in 11 comparisons and significantly less than one in 5 comparisons, with four of them happening after the year 2000. Such patterns suggest that the firms of the US P-L industry managed to produce closer to the production frontier before 2000 and experienced significant improvement in the production frontier after 2000.

Figures 6.8, 6.9, and 6.10 show the distribution of TFP, technical change, and technical efficiency change by score ranges based on 2010–2011 adjacent year Malmquist indices. Figure 6.8 shows that 54 firms experienced productivity deterioration, 144 firms have maintained their productivity, and 462 firms experienced productivity improvements, with a majority of them having a TFP score between 1.1 and 1.2. Based on Fig. 6.9, only 27 firms suffered technical deterioration; and 570 firms have enjoyed technical improvements, with majority of them having a technical change score between 1.1 and 1.2. However, when examining the technical efficiency change (Fig. 6.10), we find that only 245 firms moved closer to their production frontier in 2011 compared with 2010 (Malmquist indices of 1.1 or greater), 190 firms experienced no improvement (Malmquist indices equal to 1.0), and 225 firms suffered a technical efficiency loss (Malmquist indices less than 1.0). The results highlight an important issue for US P-L insurers: with the rapid improvement in technology, what is the best approach for firms to effectively employ new technology in such a way as to keep up with industry peers and to stay competitive?

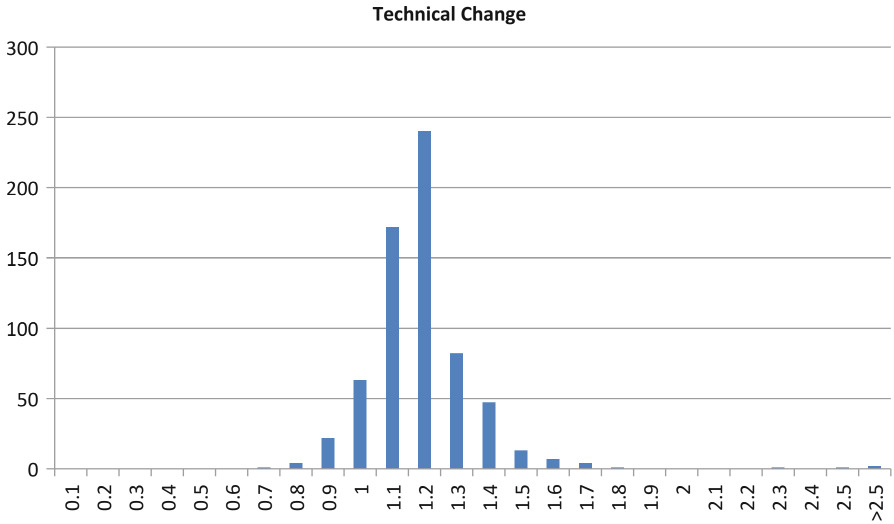


Fig. 6.9 Technical change: US P-L insurers, frequency by score, 2010–2011

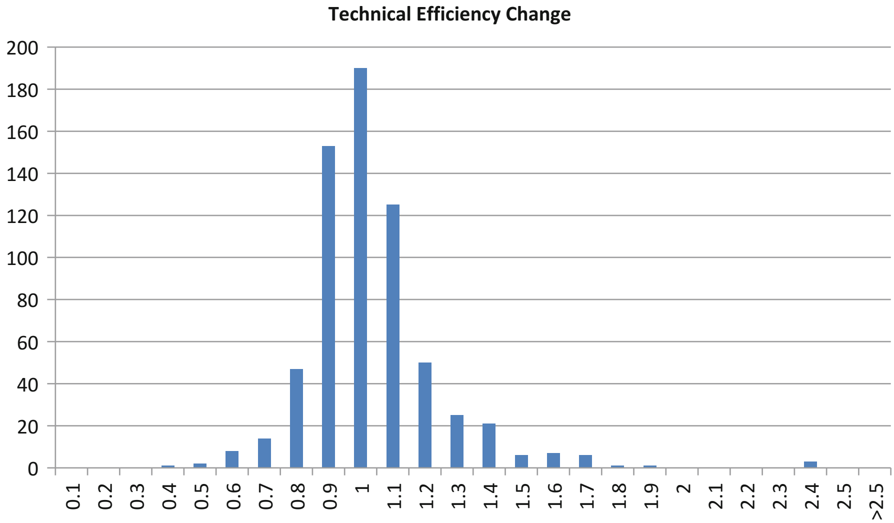


Fig. 6.10 Technical efficiency change: US P-L insurers, frequency by score, 2010–2011

6.3.4 Economies of Scale of US P-L Insurance Firms

As microeconomic theory indicates, one objective of firms is to operate with constant returns to scale (CRS) in order to minimize costs and maximize revenues. Economies of scale are present if average costs per unit of output decline as the volume of output increases. This could be achieved by spreading the firm's fixed

Table 6.3 Returns to scale of firms in the US P-L insurance industry, 1993–2011

	IRS		CRS		DRS	
Year	Number	% of total	Number	% of total	Number	% of total
1993	378	47.0	73	9.1	354	44.0
1994	369	42.9	80	9.3	411	47.8
1995	376	43.5	79	9.1	409	47.3
1996	315	37.1	92	10.8	443	52.1
1997	346	41.9	83	10.1	397	48.1
1998	364	45.1	97	12.0	346	42.9
1999	401	52.0	93	12.1	277	35.9
2000	332	45.5	94	12.9	304	41.6
2001	303	43.3	79	11.3	318	45.4
2002	395	55.1	66	9.2	256	35.7
2003	283	40.1	73	10.3	350	49.6
2004	335	46.1	69	9.5	322	44.4
2005	312	42.7	81	11.1	338	46.2
2006	393	51.9	73	9.6	292	38.5
2007	367	47.2	81	10.4	330	42.4
2008	292	37.9	74	9.6	405	52.5
2009	248	33.4	65	8.8	429	57.8
2010	290	38.9	78	10.5	377	50.6
2011	251	35.6	62	8.8	392	55.6
Mean	334	43.5	79	10.2	355	46.2

Note: *IRS* increasing returns to scale, *CRS* constant returns to scale, *DRS* decreasing returns to scale

costs over a larger volume of output. It can also be achieved if operating on a larger scale permits managers to become more specialized and therefore more efficient in carrying out specific tasks. Operating at larger scale can reduce the firm's cost of capital if income volatility is inversely related to size. This source of scale economies may be especially important in the insurance industry due to the risk-reducing impact of the law of large numbers in insurance risk pools.

Table 6.3 presents the returns to scale of US P-L insurers during the period 1993–2011. The table gives the number and percentage of firms operating with IRS (increasing returns to scale), CRS (constant returns to scale), and DRS (decreasing returns to scale). The percentage of CRS firms averages 10.2 % for the sample period as a whole. On average 43.5 % of the firms in the industry operate with IRS and 46.3 % of firms operate with DRS.

Estimation of linear time trend regressions with the percentage of firms with IRS, CRS, and DRS as dependent variables shows a statistically insignificant upward trend in the percentage of firms with DRS, which is primarily driven by the data after 2007, and a downward trend in the percentage of firms with CRS and IRS over the sample period. In general, the results suggest that it is not easy to attain full scale efficiency in this industry. The fairly high percentage of firms with DRS

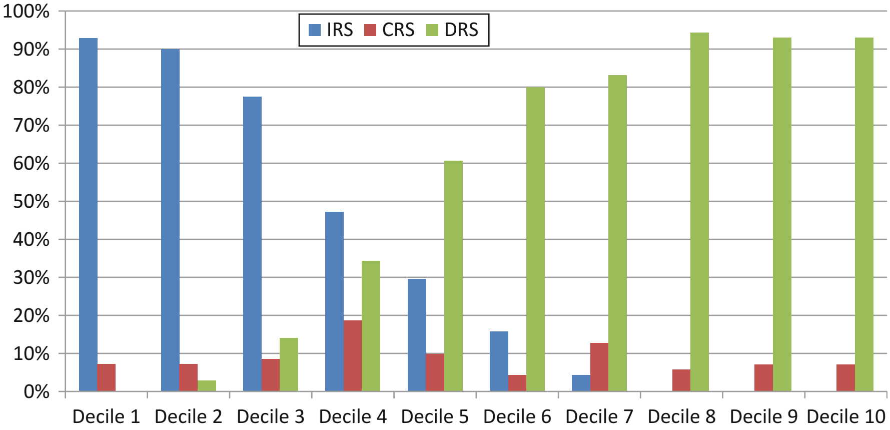


Fig. 6.11 Returns to scale by asset decile, year 2011

indicates that justifying further consolidation on the grounds of scale efficiency should be viewed with considerable skepticism.

After observing that few firms in the industry operate with CRS, it becomes interesting to study what types of firms operate with IRS and what types of firms operate with DRS. Figure 6.11 plots the proportion of firms in each asset size decile operating with IRS, CRS, and DRS in 2011. The figure indicates clearly the relationship between returns to scale and firm size. In the three smallest asset size deciles, the majority of firms operate with IRS, and the proportion of firms with IRS declines monotonically with the size deciles up to decile eight. The IRS firms should be able to reduce their costs by becoming larger, either through quick organic growth or M&As. In the six largest asset size deciles, the majority of firms operate with DRS, and the percentage of firms with DRS rises monotonically through deciles one through eight. However, a small percentage of firms in each size decile managed to operate with CRS. Thus, it is possible for even the largest and smallest firms to attain CRS, indicating that there may be important managerial lessons to learn from scale efficient insurers.

6.3.5 Efficiency by Ownership Structure

6.3.5.1 Efficiency by Organizational Form

Organizational form is often considered to be an important determinant of firm performance in the insurance industry. In the US, the two dominant forms of organization are the mutual ownership form (mutuals and reciprocals) and the stock ownership form. Two principal hypotheses have been developed regarding the relative success of mutual and stock insurers: the *expense preference hypothesis* and the *managerial discretion hypothesis*. Both hypotheses are based on the

Table 6.4 Frequency and market share of mutual vs. stock firms, 1993–2011

Year	Number of firms				Market share by NPW		Market share by asset	
	Stock		Mutual					
	N	%N (%)	N	%N (%)	Stock (%)	Mutual (%)	Stock (%)	Mutual (%)
1993	478	59.4	327	40.6	60.9	39.1	69.3	30.7
1994	514	59.8	346	40.2	62.4	37.6	72.3	27.7
1995	514	59.5	350	40.5	62.0	38.0	71.9	28.1
1996	518	60.9	332	39.1	61.7	38.3	71.1	28.9
1997	493	59.7	333	40.3	60.2	39.8	69.1	30.9
1998	468	58.0	339	42.0	59.1	40.9	67.1	32.9
1999	445	57.7	326	42.3	62.8	37.2	69.1	30.9
2000	400	54.8	330	45.2	63.5	36.5	69.1	30.9
2001	377	53.9	323	46.1	66.1	33.9	74.0	26.0
2002	389	54.3	328	45.7	64.4	35.6	72.9	27.1
2003	398	56.4	308	43.6	66.6	33.4	74.6	25.4
2004	423	58.3	303	41.7	68.9	31.1	75.9	24.1
2005	434	59.4	297	40.6	68.6	31.4	75.3	24.7
2006	452	59.6	306	40.4	67.5	32.5	74.0	26.0
2007	465	59.8	313	40.2	67.6	32.4	72.2	27.8
2008	456	59.1	315	40.9	67.0	33.0	72.2	27.8
2009	433	58.4	309	41.6	68.1	31.9	74.4	25.6
2010	437	58.7	308	41.3	64.5	35.5	71.2	28.8
2011	404	57.3	301	42.7	65.9	34.1	72.4	27.6
Mean	447	58.1	321	41.9	64.6	35.4	72.0	28.0

economic theory of agency (see Cummins et al. 1999b; Mayers and Smith 1988), which implies that the stock ownership form provides better mechanisms for owners to control the firm's managers than the mutual ownership form. To the extent that poor control mechanisms lead to sub-optimal performance, the *expense preference hypothesis* predicts that mutuals will perform more poorly than stocks. The *managerial discretion hypothesis*, predicts that mutuals will focus on less complex and less risky lines of business where managers need less discretion, and therefore owner control over managers is less important. If the expense preference hypothesis predominates, mutuals are expected to perform more poorly than stocks; but if the managerial discretion hypothesis predominates, there should be no significant differences in performance between mutuals and stocks.

Table 6.4 presents the numbers and market share of mutual and stock insurers by net premiums written and by total admitted assets during our sample period. On average, about 41.9 % of DMUs in the industry are mutual companies, and 58.1 % are stock companies. The percentage of mutual firms in the market does not change much over time. When looking at the market share of the two types of ownership, we find that, on average, mutual firms account for 35.4 % of total premiums written in the market, with an asset percentage of 28 %. Stock firms collect 64.6 % of the total premiums written and manage 72 % of the total assets of the industry.

Estimation of linear time trend regressions with the percentage of firms of mutual and stock shows that we have a relatively stable combination of mutual and stock firms in the market over time. However, the time trend regressions with market shares by net premiums written and assets as dependent variables show statistically significant upward trends in the market share of stock firms and significant downward trends in the market share of mutual firms over the sample period. In general, the results suggest stock firms do have some advantages over mutuals.

Table 6.5 presents the efficiency of mutual and stock firms during our sample period. The results show that mutual firms are consistently more cost- and allocatively efficient than stock firms in most years of the sample, but they are consistently less revenue efficient than stocks, which leads to insignificant differences in profit inefficiency in the majority of years. The tests based on the all-year average are consistent with the observed pattern: mutuals are more cost- and allocatively efficient but less revenue efficient and show no difference from stocks in technical or profit efficiency. Noticeably, this "general" pattern may change in the future. In 2011, mutual insurers demonstrated significantly lower technical efficiency than stock insurers, offsetting the allocative efficiency advantage of mutual firms. No significant cost efficiency advantage combined with the disadvantage in revenue efficiency causes mutual insurers to be less profit efficient than stock insurers in 2011. These differences are illustrated in Fig. 6.12.

6.3.5.2 Efficiency by Ownership Status of Firm

In this chapter, we also examine another type of ownership variation in the US P-L insurance market: whether publicly traded P-L insurers perform better or worse than non-traded firms in terms of frontier efficiency. Literature on corporate finance has provided plentiful arguments regarding the costs and benefits of public listing. The benefits of operating as a public company are many: increasing a firm's ability to raise equity capital and overcome capital constraints, creating more effective managerial discipline (strong external monitoring mechanisms), facilitating M&A activities, and enhancing the reputation of the firm (Brau and Fawcett 2006; Xie 2010). However, there are additional costs and challenges for public firms. Publicly traded firms have to comply with federal securities laws and filing requirements. The burden of filing became even greater after the passage of the Dodd-Frank Act in 2010. Public firms may also lose flexibility because they are subject to scrutiny from public shareholders and have fiduciary duties to public shareholders. As shown by Brau and Fawcett (2006), the desire to maintain decision-making control is a CFO's most important motivation for staying private. If the benefits of public trading outweigh the costs, we should observe better performance of public firms; otherwise, public firms' performance may suffer as the costs outweigh the benefits.

Table 6.6 presents the numbers and percentages of public and private insurers over time and their market share by net premiums written and admitted assets. On average, only 11.9 % of insurers are publicly traded, with the remaining staying

Table 6.5 Efficiency of mutual vs. stock firms, 1993–2011

	Cost efficiency				Allocative efficiency				Technical efficiency				Revenue efficiency				Profit inefficiency			
Year	S	M	M-S	*T*-test	S	M	M-S	*T*-test	S	M	M-S	*T*-test	S	M	M-S	*T*-test	S	M	M-S	*T*-test
1993	0.43	0.49	0.06	***	0.71	0.75	0.05	***	0.62	0.66	0.04	**	0.42	0.36	−0.07	***	0.79	0.70	−0.09	**
1994	0.45	0.50	0.05	***	0.71	0.78	0.06	***	0.64	0.65	0.01		0.48	0.41	−0.07	***	0.62	0.59	−0.03	
1995	0.46	0.54	0.08	***	0.71	0.81	0.10	***	0.64	0.67	0.03	**	0.46	0.42	−0.04	***	0.61	0.55	−0.07	**
1996	0.47	0.53	0.06	***	0.70	0.79	0.10	***	0.68	0.68	0.00		0.50	0.45	−0.04	***	0.57	0.54	−0.03	
1997	0.42	0.48	0.06	***	0.66	0.76	0.10	***	0.65	0.64	0.00		0.41	0.33	−0.08	***	0.92	1.09	0.18	***
1998	0.51	0.57	0.05	***	0.77	0.84	0.07	***	0.68	0.68	0.01		0.47	0.41	−0.07	***	0.59	0.55	−0.04	
1999	0.53	0.60	0.06	***	0.78	0.86	0.08	***	0.69	0.70	0.01		0.49	0.42	−0.07	***	0.52	0.49	−0.04	
2000	0.53	0.58	0.05	***	0.76	0.82	0.06	***	0.70	0.71	0.01		0.50	0.43	−0.07	***	0.53	0.52	−0.01	
2001	0.51	0.58	0.06	***	0.77	0.84	0.07	***	0.67	0.69	0.02		0.44	0.39	−0.05	***	0.67	0.63	−0.04	
2002	0.48	0.48	0.01		0.75	0.79	0.03	***	0.64	0.63	−0.01		0.44	0.39	−0.06	***	0.72	0.76	0.04	
2003	0.50	0.53	0.03	**	0.76	0.81	0.05	***	0.66	0.66	0.00		0.45	0.43	−0.02		0.67	0.63	−0.04	
2004	0.52	0.55	0.03	***	0.78	0.82	0.03	***	0.66	0.68	0.01		0.45	0.41	−0.04	***	0.67	0.69	0.01	
2005	0.52	0.55	0.03	**	0.77	0.82	0.05	***	0.69	0.68	−0.01		0.50	0.44	−0.06	***	0.51	0.57	0.06	*
2006	0.47	0.51	0.04	***	0.73	0.79	0.06	***	0.65	0.64	−0.01		0.44	0.40	−0.04	***	0.67	0.67	0.00	
2007	0.48	0.53	0.05	***	0.75	0.84	0.08	***	0.64	0.64	−0.01		0.44	0.39	−0.05	***	0.71	0.69	−0.01	
2008	0.48	0.51	0.03	***	0.75	0.82	0.07	***	0.64	0.62	−0.01		0.41	0.35	−0.07	***	0.74	0.75	0.02	
2009	0.52	0.54	0.02	*	0.80	0.86	0.05	***	0.66	0.64	−0.01		0.50	0.46	−0.04	***	0.49	0.50	0.02	
2010	0.54	0.57	0.03	**	0.82	0.87	0.06	***	0.67	0.66	−0.01		0.52	0.48	−0.04	***	0.40	0.40	0.01	
2011	0.50	0.51	0.01		0.76	0.84	0.08	***	0.66	0.61	−0.05	***	0.48	0.41	−0.06	***	0.53	0.60	0.06	**
All Years	0.49	0.53	0.04	***	0.75	0.81	0.07	***	0.66	0.66	0.00		0.46	0.41	−0.05	***	0.63	0.63	0.00	

Note: *M* mutual insurers, *S* stock insurers. *Significant at the 10 % level; **significant at the 5 % level; ***significant at the 1 % level

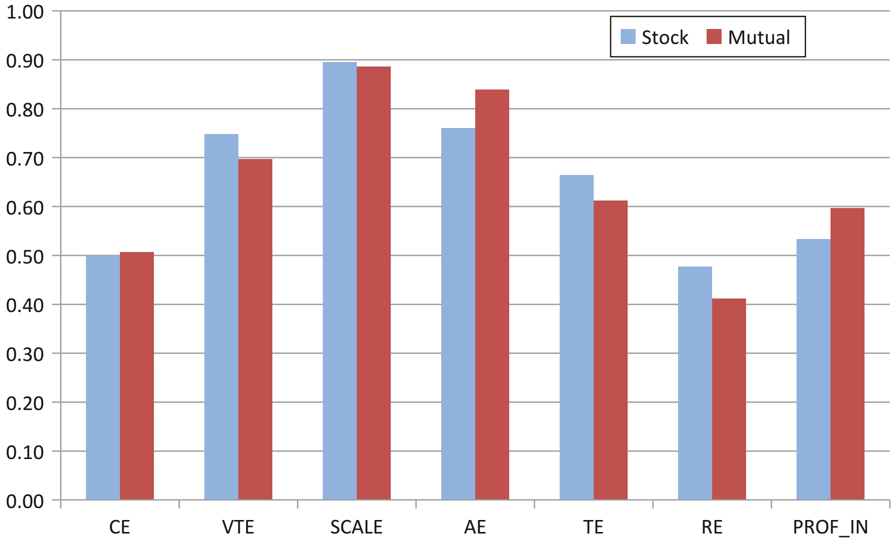


Fig. 6.12 Mean efficiency of mutual vs. stock firms, year 2011

Table 6.6 Frequency and market share of public vs. private firms, 1993–2011

	Number of firms				Market share by NPW		Market share by asset	
	Public		Private		Public	Private	Public	Private
Year	N	%N (%)	N	%N (%)	(%)	(%)	(%)	(%)
1993	92	11.4	713	88.6	38.4	61.6	44.2	55.8
1994	98	11.4	762	88.6	40.2	59.8	48.7	51.3
1995	105	12.2	759	87.8	46.2	53.8	54.5	45.5
1996	111	13.1	739	86.9	46.9	53.1	54.4	45.6
1997	111	13.4	715	86.6	48.9	51.1	56.7	43.3
1998	113	14.0	694	86.0	50.7	49.3	57.5	42.5
1999	99	12.8	672	87.2	53.1	46.9	59.4	40.6
2000	91	12.5	639	87.5	56.6	43.4	61.8	38.2
2001	84	12.0	616	88.0	55.9	44.1	61.4	38.6
2002	89	12.4	628	87.6	53.8	46.2	58.8	41.2
2003	86	12.2	620	87.8	55.3	44.7	61.7	38.3
2004	90	12.4	636	87.6	57.1	42.9	62.7	37.3
2005	94	12.9	637	87.1	57.6	42.4	63.3	36.7
2006	95	12.5	663	87.5	55.9	44.1	61.2	38.8
2007	92	11.8	686	88.2	55.5	44.5	59.1	40.9
2008	83	10.8	688	89.2	49.9	50.1	55.9	44.1
2009	74	10.0	668	90.0	47.2	52.8	52.6	47.4
2010	71	9.5	674	90.5	46.5	53.5	54.1	45.9
2011	63	8.9	642	91.1	47.5	52.5	53.7	46.3
Mean	92	11.9	676	88.1	50.7	49.3	56.9	43.1

private. However, this 19.5 % have underwritten 50.7 % of the premiums in the market and accounted for 56.9 % of industry assets. The market share of public firms has increased over time based on a comparison of 1993 and 2011. However, the linear time trends in public firm shares of net premiums written and assets are not statistically significant.

The efficiency scores of public and private firms are presented in Table 6.7. We find that before 1999, public firms in general were less revenue efficient than private firms; have no advantage in cost efficiency, allocative efficiency or technical efficiency; and are therefore less profit efficient. The profit efficiency discrepancy between public and private firms is not statistically significant after 1999. The sample average efficiency score shows that public firms are more cost efficient but less revenue efficient than private firms, such that profit efficiency is not statistically different between public and private firms.

Figure 6.13 presents the efficiency of public and private P-L insurers in 2011. Public firms are significantly more cost and revenue efficient than private firms in 2011, but these differences do not translate into a statistically significant advantage in profit efficiency.

6.3.6 Efficiency by Product Line

The US P-L insurance industry distributes a very broad range of products to the market. Personal lines products such as auto insurance and homeowners insurance are less complex than commercial lines and also involve fewer service components (e.g., loss control and claims adjusting services) in the product. The risk pooling and law of large number rules work perfectly for these personal lines. Commercial lines, on the other hand, often involve more complex risks, are more demanding in terms of underwriting and loss control services, and are more costly in claims adjustment. In addition, some commercial lines have relatively short claims payout tails, while others (especially liability-related lines) have much longer payout tails, such that pricing is more sensitive to interest changes and costs are more difficult to predict. Complex products also tend to be more expensive to produce, particularly those requiring more resources in underwriting, sales, and services. On the other hand, these products might also generate higher profit margins for insurers. As a result, it is likely that choice of product offerings can result in differences in efficiency across insurers. In this section, we analyze the relationship between firm efficiency and its specialization (diversification) in product range.

We classify the product offerings of insurers into seven categories: personal lines short-tail only, personal lines long-tail only, commercial lines short-tail only, commercial lines long-tail only, personal lines only (have business in both short-tail and long-tail personal lines), commercial lines only (have business in both short-tail and long-tail commercial lines), and both commercial and personal lines. Table 6.8 presents the frequency of firms by breadth of product range during 1993–2011. We find that very few firms choose to offer only short-tail personal lines or long-tail

Table 6.7 Efficiency of public vs. private firms, 1993–2011

	Cost efficiency				Allocative efficiency				Technical efficiency				Revenue efficiency				Profit inefficiency			
Year	Pu	Pr	Pu-Pr	*T*-test	Pu	Pr	Pu-Pr	*T*-test	Pu	Pr	Pu-Pr	*T*-test	Pu	Pr	Pu-Pr	*T*-test	Pu	Pr	Pu-Pr	*T*-test
1993	0.45	0.45	0.00		0.75	0.72	0.03	*	0.61	0.64	−0.03		0.39	0.40	−0.01		0.91	0.74	0.17	**
1994	0.46	0.47	−0.01		0.75	0.74	0.01		0.62	0.65	−0.03	*	0.41	0.46	−0.04	**	0.72	0.59	0.13	**
1995	0.47	0.49	−0.02		0.74	0.75	−0.01		0.64	0.66	−0.02		0.42	0.45	−0.03	**	0.68	0.57	0.10	**
1996	0.49	0.49	0.00		0.72	0.74	−0.02		0.68	0.68	0.01		0.45	0.48	−0.03	*	0.64	0.55	0.09	**
1997	0.44	0.44	0.00		0.70	0.69	0.00		0.64	0.65	−0.01		0.34	0.39	−0.05	**	1.08	0.97	0.11	
1998	0.56	0.53	0.03		0.81	0.79	0.02		0.69	0.68	0.01		0.42	0.45	−0.03	*	0.66	0.56	0.10	*
1999	0.59	0.56	0.03	*	0.84	0.81	0.03	**	0.70	0.69	0.01		0.45	0.46	−0.01		0.52	0.50	0.02	
2000	0.59	0.54	0.04	**	0.82	0.78	0.04	***	0.72	0.70	0.02		0.46	0.47	−0.01		0.54	0.52	0.02	
2001	0.58	0.54	0.04	**	0.85	0.79	0.06	***	0.68	0.68	0.00		0.41	0.42	−0.01		0.67	0.65	0.02	
2002	0.52	0.47	0.05	***	0.84	0.76	0.08	***	0.63	0.64	−0.01		0.39	0.42	−0.03		0.74	0.74	0.00	
2003	0.53	0.51	0.03		0.82	0.77	0.04	***	0.66	0.66	0.01		0.44	0.45	−0.01		0.63	0.66	−0.03	
2004	0.58	0.52	0.06	***	0.84	0.79	0.06	***	0.69	0.67	0.02		0.42	0.44	−0.02		0.71	0.68	0.04	
2005	0.57	0.53	0.05	***	0.83	0.78	0.05	***	0.70	0.68	0.01		0.47	0.48	−0.01		0.50	0.54	−0.04	
2006	0.53	0.48	0.05	***	0.79	0.75	0.04	***	0.67	0.65	0.02		0.43	0.42	0.00		0.63	0.68	−0.05	
2007	0.52	0.50	0.03	*	0.84	0.78	0.06	***	0.63	0.64	−0.01		0.41	0.42	−0.01		0.74	0.69	0.05	
2008	0.51	0.49	0.02		0.79	0.78	0.01		0.64	0.63	0.01		0.38	0.39	−0.01		0.78	0.74	0.05	
2009	0.57	0.53	0.04	**	0.85	0.82	0.03	*	0.67	0.65	0.02		0.51	0.48	0.03		0.49	0.49	−0.01	
2010	0.58	0.55	0.03		0.88	0.84	0.04	***	0.67	0.67	0.00		0.52	0.50	0.02		0.39	0.40	−0.01	
2011	0.56	0.50	0.07	***	0.82	0.79	0.03	*	0.68	0.64	0.04	**	0.50	0.44	0.05	**	0.54	0.56	−0.02	
All Years	0.53	0.50	0.03	***	0.80	0.77	0.03	***	0.66	0.66	0.00		0.43	0.44	−0.01	***	0.67	0.62	0.05	

Note: *Pu* publicly traded insurers, *Pr* private insurers. *Significant at the 10 % level; **significant at the 5 % level; ***significant at the 1 % level

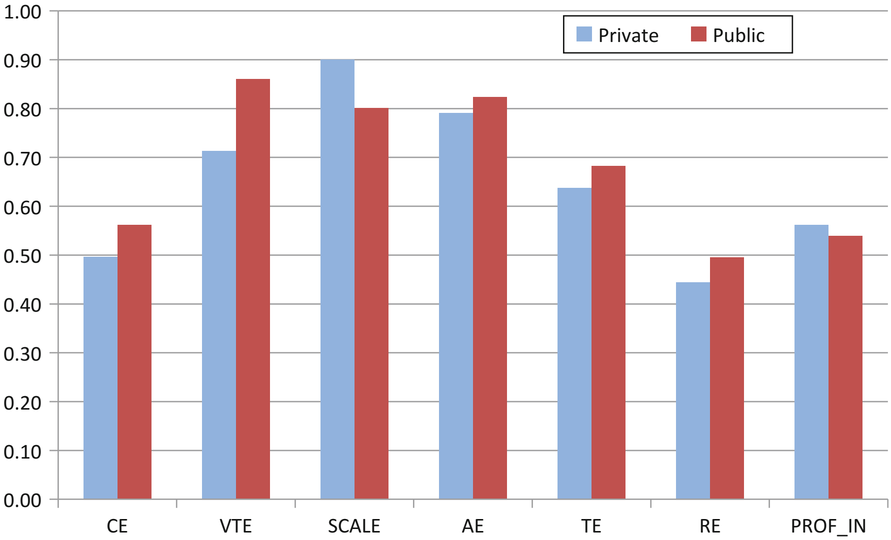


Fig. 6.13 Mean efficiency of public vs. private firms, year 2011

Table 6.8 Frequency of firms by breadth of product range, 1993–2011

Year	Total firms	% PST only	% PLT only	% CST only	% CLT only	% Personal lines only	% Commercial lines only	% Both personal and commercial	
								All	Personal >75 %
1993	805	0.4	0.0	9.3	5.1	2.6	8.2	74.4	21.7
1994	860	0.3	0.0	11.0	5.8	3.0	7.3	72.4	19.3
1995	864	0.6	0.0	9.1	10.3	2.3	9.3	68.4	19.3
1996	850	0.4	0.1	10.7	10.4	2.5	9.8	66.2	19.9
1997	826	0.4	0.2	10.3	10.2	2.7	10.8	65.5	20.1
1998	807	0.4	0.1	11.2	10.4	2.4	9.2	66.4	20.7
1999	771	0.3	0.3	10.5	10.8	3.1	10.5	64.6	20.9
2000	730	0.3	0.4	11.1	11.5	2.6	11.1	63.0	20.0
2001	700	0.3	0.1	10.6	12.3	3.0	9.0	64.7	20.7
2002	717	0.3	0.3	10.0	12.8	2.6	10.0	63.9	21.8
2003	706	0.3	0.0	10.1	14.4	2.8	8.6	63.7	20.5
2004	726	0.3	0.3	9.5	15.2	3.7	10.1	61.0	19.3
2005	731	0.1	0.3	8.9	15.5	3.4	10.8	61.0	20.4
2006	758	0.1	0.7	8.4	17.0	4.1	11.6	58.0	19.9
2007	778	0.1	0.9	9.1	17.5	3.3	11.4	57.6	19.7
2008	771	0.1	0.6	7.7	16.7	3.4	12.7	58.8	20.4
2009	742	0.1	0.5	8.0	17.3	3.6	11.2	59.3	19.5
2010	745	0.0	0.5	8.5	16.8	3.0	12.1	59.2	21.1
2011	705	0.0	0.3	8.2	17.7	3.8	11.6	58.3	21.0

Note: *PST* personal lines short-tail, *PLT* personal lines long-tail, *CST* commercial lines short-tail, *CLT* commercial lines long-tail

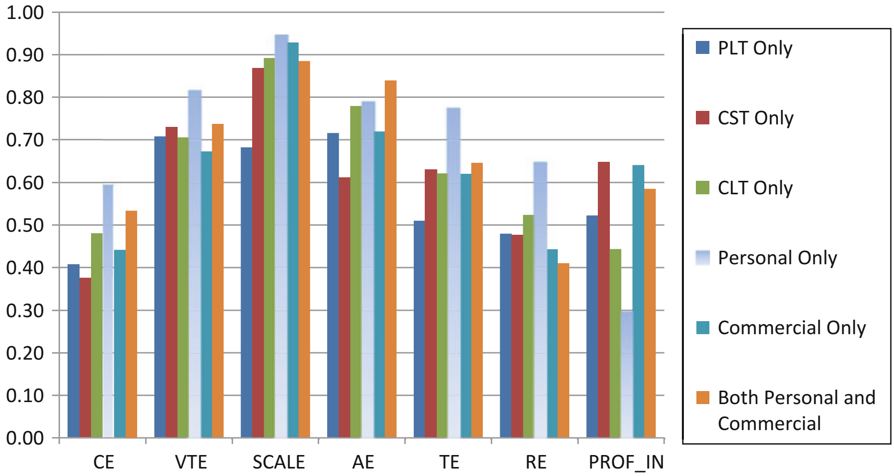


Fig. 6.14 Efficiency by product diversification, year 2011

personal lines, or only personal lines. Fewer than 10 % of firms in the industry choose to specialize in short-tail commercial lines only, but more firms (especially in the years after 2000) specialize in long-tail commercial lines only. There are another 10 % of firms choosing to offer both short-tail and long-tail commercial lines but no personal lines. The majority of firms (an average of 63 %) in the industry offer both personal lines and commercial lines, and about 30 % of these firms have more than 75 % of premium income from personal lines. There is a statistically significant downward time trend for our time period in the number of firms offering both personal and commercial lines. Thus, more insurers have chosen to specialize.

Figure 6.14 shows insurer efficiency by product lines based for 2011. We find that firms specializing in personal lines only have the highest efficiency score in all types of efficiency except for allocative efficiency. Firms specializing in long-tail commercial lines rank second in revenue efficiency and profit efficiency, while firms diversifying across both personal lines and commercial lines rank second in cost efficiency.

We further divide firms that offer both personal and commercial lines into those that are predominantly personal lines firms and those that are predominantly commercial lines firms. Firms are classified as predominantly personal lines if the percentage of personal lines premiums written is at least 75 % of total premiums written. Figure 6.15 presents the efficiency comparison of these two types of firms. We find that firms with personal lines predominating are more cost efficient (with higher technical efficiency, pure technical efficiency, and scale efficiency) and are therefore more profit efficient. This is expected as size and automated systems are more advantageous in personal lines.

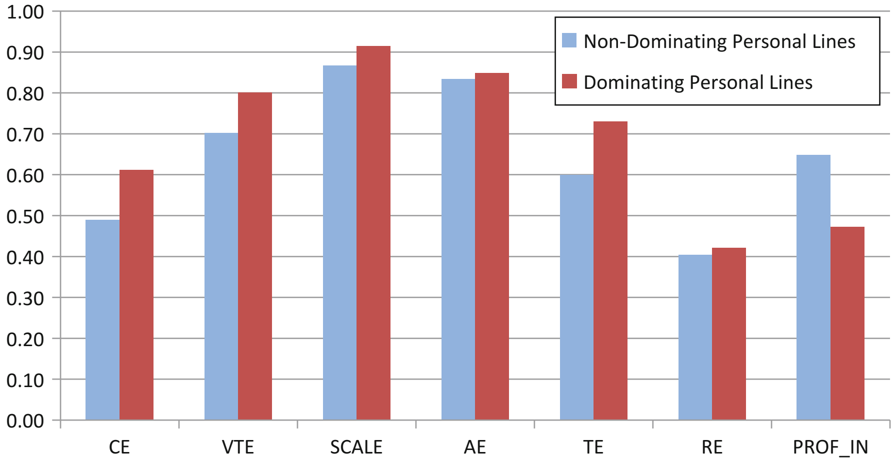


Fig. 6.15 Efficiency by product diversification, dominating personal lines vs. others, year 2011

6.3.7 Efficiency by Distribution System

The relationship between distribution systems and efficiency in the insurance industry has been analyzed extensively in the prior literature. This is a natural topic for the insurance industry because multiple distribution systems have long coexisted in the industry. According to A.M. Best's classification, we classify the insurance distribution system into agency based channel (which includes independent agents, general agents, managing general agents, career agents, other agencies, and banks), the brokerage channel, direct writers (which includes internet, exclusive/captive agents, direct response, affinity group marketing, worksite marketing, and other direct), and combined channels (mixture of different channels mentioned above). We single out the brokerage channel from the "agency based" category because brokers are often larger and are more likely to focus on commercial lines of insurance for medium to large-scale buyers.

Table 6.9 presents the frequency and market share of US P-L insurers by distribution system (1993–2011). A majority of firms in the industry are still using agency based distribution channels; however, the percentage of firms using pure agency based channel has decreased significantly over time, as do pure brokerage based and direct writers. Meanwhile, an increasing percentage of firms have adopted combined distribution systems to maximize business sales.

Regarding market share, there is a statistically significant downward linear time trend in the market share of direct marketers, statistically significant upward trends in the market shares of insurers using brokers or mixed distribution systems, and no significant trend in the market share of insurers primarily using the agency distribution channel.

Table 6.9 Frequency and market share of the US P-L insurers by distribution system, 1993–2011

	Number of firms								Market share by NPW				Market share by asset			
	Agent		Broker		Direct		Mix									
Year	N	%N (%)	N	%N (%)	N	%N (%)	N	%N (%)	Agent (%)	Broker (%)	Direct (%)	Mix (%)	Agent (%)	Broker (%)	Direct (%)	Mix (%)
1993	549	68.2	57	7.1	152	18.9	42	5.2	45.8	5.7	44.5	3.9	52.7	7.5	36.3	3.5
1994	533	62.0	66	7.7	154	17.9	42	4.9	45.0	5.9	43.5	3.7	47.9	7.6	33.6	3.2
1995	532	61.6	61	7.1	157	18.2	46	5.3	42.8	6.5	44.5	2.5	44.1	7.8	34.5	2.7
1996	526	61.9	59	6.9	164	19.3	52	6.1	42.8	7.1	44.1	2.6	44.5	8.6	35.1	2.9
1997	504	61.0	57	6.9	154	18.6	61	7.4	41.7	6.8	45.4	2.8	43.7	8.2	36.8	2.7
1998	483	59.9	53	6.6	155	19.2	56	6.9	41.6	6.4	46.1	2.7	43.9	7.8	37.1	2.8
1999	440	57.1	43	5.6	137	17.8	52	6.7	43.2	7.1	43.1	3.0	44.7	8.3	34.8	3.1
2000	405	55.5	36	4.9	127	17.4	57	7.8	40.6	6.8	42.9	5.3	42.4	7.9	35.7	3.9
2001	387	55.3	38	5.4	116	16.6	55	7.9	43.9	7.7	41.6	6.1	51.9	9.0	34.0	4.8
2002	393	54.8	40	5.6	113	15.8	70	9.8	41.2	9.5	40.9	7.3	50.6	9.4	33.0	5.7
2003	413	58.5	41	5.8	133	18.8	73	10.3	42.3	11.0	38.5	8.0	45.7	10.9	32.0	6.0
2004	425	58.5	44	6.1	125	17.2	78	10.7	46.6	11.7	36.3	5.1	48.1	11.2	30.1	4.9
2005	424	58.0	43	5.9	120	16.4	82	11.2	47.4	10.9	35.6	5.0	49.0	11.5	28.6	5.2
2006	424	55.9	39	5.1	126	16.6	89	11.7	45.0	11.6	36.9	5.3	46.6	11.9	30.9	5.5
2007	426	54.8	35	4.5	125	16.1	100	12.9	45.1	10.6	37.0	6.7	49.0	12.1	31.4	7.0
2008	421	54.6	40	5.2	122	15.8	104	13.5	43.0	10.8	38.6	6.9	47.2	13.5	31.1	7.6
2009	378	50.9	28	3.8	104	14.0	100	13.5	45.1	8.5	38.5	7.2	49.9	11.7	30.1	7.6
2010	366	49.1	31	4.2	106	14.2	97	13.0	40.6	8.6	41.8	7.9	46.1	12.2	32.1	8.5
2011	350	49.6	29	4.1	99	14.0	92	13.0	42.5	8.7	39.7	8.0	47.6	11.9	29.7	9.9
Mean	441	57.2	44	5.7	131	17.0	71	9.4	43.5	8.5	41.0	5.3	47.1	9.9	33.0	5.1

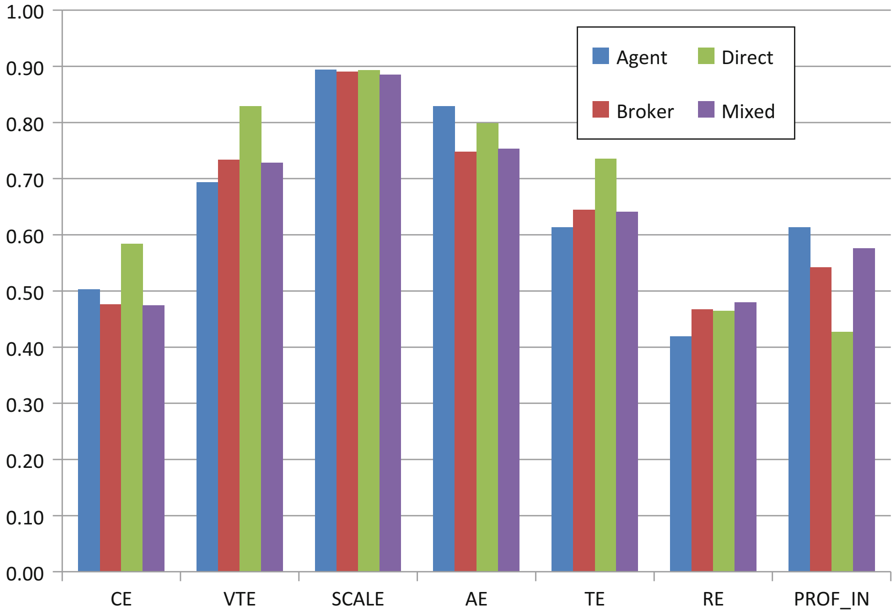


Fig. 6.16 Efficiency by distribution system, year 2011

The traditional finding in the P-L insurance industry has been that independent distributors tend to have higher costs than direct writers. However, Berger et al. (1997) provide evidence that, although independent distributors tend to be less cost efficient, the higher costs represent the provision of additional services that are valued by buyers such that independent and direct distributors are roughly equal in revenue and profit efficiency. We plot the efficiency of insurers with various distribution channels in Fig. 6.16 based on 2011 data. The figure shows that direct writers are the most cost efficient (due to higher pure technical efficiency) and profit efficient. Agency based firms are the most allocatively efficient, while firms adopting mixed distribution channels are the most revenue efficient.

The finding that direct writers are more cost efficient is consistent with the existing literature. Firms using the direct writing (vertically integrated) distribution system are predicted to be more efficient to the extent that such firms can more easily recognize cost savings through automation and more fully realize the benefits of technology investments in customer databases and marketing than can insurers using less vertically integrated distribution systems (Carr et al. 1999). The higher technical efficiency scores of direct writers provide support for the argument that vertical integration is associated with higher efficiency because of the ability to better employ automation in reducing costs and increasing revenues.

The finding that agency based insurers demonstrate higher allocative efficiency than firms using direct marketing, brokerage, and mixed distribution systems perhaps suggests that more resource allocation is conducted at the agency level in the case of independent agents.[12] The higher revenue efficiency of the mixed distribution system helps explain the increased number of firms that switch to this channel. However, the results shown here cannot adequately explain the coexistence of multiple distributional channels. We find that an agency based distribution system is clearly inferior from both a cost and revenue perspective. Further studies on this issue are necessary.

6.3.8 Regression Analyses on Determinants of Efficiency

The last step in the analysis is to examine the factors that affect insurers' efficiency and productivity and explain the variations in efficiency scores. In this section, we analyze the relationship between firm characteristics and efficiency scores by conducting a panel data regression analysis on the firms in the industry. The dependent variables in the regressions are efficiency scores, and the independent variables are firm characteristics. The regressions are conducted using unbalanced panel data in order to maximize the number of firms included in the analysis and avoid any problems with survivor bias that might be inherent in the use of a balanced panel.

We run regressions with one-way time fixed effects only and also run regressions with two-way fixed effects models (both firm and time fixed effects controlled). The regression model for a two-way design is summarized as follows:

$$y_{it} = \alpha + \beta' x_{it} + \varepsilon_{it} + u_i + w_t,$$

where i indexes firms (DMUs) and t indexes the time periods. The dependent variable y_{it} is firm i's efficiency in year t, x_{it} is the set of independent variables, u_i controls for firm fixed effects, w_t controls for effects for years, and ε_{it} is the overall regression error.

The regression results are reported in Table 6.10. We conduct overall regressions for cost efficiency, revenue efficiency, profit *inefficiency*, and total factor productivity change of firms. Since a Wald test of firm fixed effects shows that the two-way fixed effects models are superior to the models without firm fixed effects, we focus most of the discussion on these results. The one-way fixed effect results are also shown for completeness.

The control variables x_{it} included in the regressions are discussed below, along with the regression results. The discussion is based on two-way fixed effects unless otherwise mentioned.

[12] Although this also might seem to be true for brokers, who are also independent distributors, the brokerage market tends to deal with larger and more complex risks than the independent agency market, potentially posing significantly greater resource allocation problems.

Table 6.10 Regression analyses of the determinants of firm efficiency and productivity

Variables	Definition of variable	Two-way fixed effects				One-way fixed effects			
		Cost efficiency	Revenue efficiency	Profit inefficiency	Total factor productivity	Cost efficiency	Revenue efficiency	Profit inefficiency	Total factor productivity
Size	ln (assets)	0.0370***	0.0161***	−0.0287***	−0.0625***	0.0299***	0.0191***	−0.0511***	−0.0101***
		[0.002]	[0.003]	[0.007]	[0.007]	[0.001]	[0.001]	[0.002]	[0.002]
Financial leverage	Liabilities/Surplus	0.0150***	0.0160***	−0.0196***	0.0089*	0.0239***	0.0206***	−0.0566***	0.0034
		[0.001]	[0.002]	[0.005]	[0.005]	[0.001]	[0.002]	[0.004]	[0.003]
Underwriting leverage	NPW/surplus	0.0379***	0.0945***	−0.1943***	−0.0314***	0.0278***	0.1002***	−0.1934***	−0.0154***
		[0.002]	[0.003]	[0.008]	[0.008]	[0.002]	[0.003]	[0.006]	[0.005]
Personal	Percentage of NPW in personal lines	0.0721***	−0.0055	−0.0793***	0.0036	0.1916***	−0.0236***	−0.1793***	0.0152**
		[0.004]	[0.006]	[0.015]	[0.028]	[0.003]	[0.004]	[0.010]	[0.008]
Pwherf	HHI across states by NPW	0.0548***	0.0901***	−0.1149***	−0.0751***	0.0759***	0.0560***	−0.1762***	−0.0106
		[0.007]	[0.010]	[0.026]	[0.025]	[0.004]	[0.005]	[0.013]	[0.008]
Lbherfpw	HHI across business lines by NPW	0.1028***	0.1472***	−0.2113***	−0.1436***	0.0507***	0.2313***	−0.3034***	−0.0116
		[0.008]	[0.011]	[0.029]	[0.028]	[0.004]	[0.006]	[0.015]	[0.010]
Direct	Direct writer dummy	0.003	0.0258***	−0.0317**	−0.0216	0.0569***	0.0644***	−0.1869***	0.0024
		[0.005]	[0.006]	[0.016]	[0.016]	[0.003]	[0.004]	[0.010]	[0.006]
Mutual	Mutual insurer dummy	0.005	−0.0087	0.0116	−0.0319	0.0200***	−0.0167***	−0.0147*	−0.0369***
		[0.006]	[0.008]	[0.021]	[0.020]	[0.002]	[0.003]	[0.008]	[0.005]
List	Publicly traded insurer dummy	0.0012	−0.0016	−0.0014	0.0104	−0.0106***	−0.0200***	0.0601***	0.0016
		[0.007]	[0.009]	[0.024]	[0.024]	[0.004]	[0.005]	[0.013]	[0.008]
Constant		−0.1667***	0.0414	1.3385***	2.4439***	−0.3404***	−0.2713***	2.4319***	1.2436***
		[0.049]	[0.067]	[0.174]	[0.171]	[0.015]	[0.021]	[0.051]	[0.034]
Firm dummies		Yes	Yes	Yes	Yes	No	No	No	No
Year dummies		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations		14,069	14,068	14,025	11,892	14,069	14,068	14,025	11,892
r2_adj		0.786	0.723	0.679	0.0674	0.47	0.347	0.316	0.0187
Wald-test of firm dummies (p-value)		0.0000	0.0000	0.0000	0.0000				

Note: Standard errors are in brackets. *Significant at the 10 % level; **significant at the 5 % level; ***significant at the 1 % level

Size: Firm size (which is measured as logarithm value of a firm's admitted assets) is frequently included in efficiency studies as a control variable. Cummins and Zi (1998) and Cummins and Nini (2002) find that larger firms tend to be more cost and revenue efficient, usually because large firms define the production technology for the industry and are more efficient in the allocation of resources. Contrary to this finding, Yuengert (1993) finds the relationship between firm efficiency and size to be statistically insignificant. Our regression results show that firm size is positively related to both cost and revenue efficiency. Consequently, large firms are found to be more profit efficient as well. However, size is inversely related to the total factor productivity change of insurers, suggesting that it is more difficult for large firms to improve productivity significantly over time.

Financial Leverage: Financial leverage, measured as total liabilities divided by surplus, can impact the efficiency of firms as well. Firms with a reasonably higher leverage ratio may be more cost efficient because higher leverage indicates a more efficient use of equity capital input, holding everything else equal. However, the relationship between leverage and revenue efficiency is less clear. While debt financing can provide relatively cheaper capital for insurers' operation, a higher debt ratio also subjects the firm to the probability of increased levels of financial distress (Mayers and Smith 1987; Cummins and Danzon 1997), implying that such firms would receive lower output prices and therefore have lower revenue efficiency than better capitalized firms. Our regression results show a positive impact of financial leverage on firm efficiency. Higher leverage is positively related to insurers' cost, revenue, profit efficiency, and total factor productivity change. This finding suggests that equity capital financing is costly, and using financial leverage enables insurers to enjoy the benefits of debt financing and therefore improve performance.

Underwriting Leverage: Underwriting leverage, measured as the ratio of net premiums written to policyholders' surplus, is often included in efficiency studies to control for differences in capitalization among insurers. This variable is predicted to be positively related to cost efficiency, because firms with higher premiums-to-surplus ratios use less capital relative to revenues. The relationship of this variable to revenue efficiency is ambiguous. Higher ratios of premiums-to-surplus could indicate an efficient use of resources, i.e., could be associated with more efficient risk management that permits firms to use less capital, predicting a positive relationship. On the other hand, higher premiums-to-surplus ratios suggest higher underwriting risk that is associated with higher insolvency, which may undermine the contract terms with policyholders and result in lower revenue efficiencies than better capitalized firms. The regressions support the efficiency interpretation in general—the premiums-to-surplus ratio is positively related to the cost, revenue, and profit efficiency of firms, suggesting that the efficient use of capital is a core competency which tends to be an important driver of performance in the insurance industry. However, the relationship is negative in the total factor productivity change regression. This implies that a firm with higher underwriting leverage has a lower capacity in growing its business, which results in slow growth in productivity.

Line of Business: In the previous section, we found that the difference in the range of products results in variation in the efficiency of firms. As a result, we include in the regression a variable that identifies the percentage of net premiums written in personal lines by each insurer (*Personal*). The two-way fixed effects regressions show that firms with higher proportions of business in personal lines are more cost and profit efficient, but there is no impact on revenue efficiency and TFP changes. This is as expected, as firms may be less likely to make mistakes in designing technologies or allocating resources in less complex personal lines.

Diversification: Existing literature has provided conflicting arguments about the impact of diversification on firm performance. On the one hand, diversification across lines and geographical areas tends to reduce the risk of insurers, and diversified firms may benefit from scope economies on the revenue side if buyers prefer one-stop shopping (Cummins et al. 2010). These predict a positive relationship between diversification and efficiency. On the other hand, diversification may increase agency costs (Mayers and Smith 1988), and there is growing literature documenting the tendency of diversified firms to perform poorly relative to firms that adopt the strategy of strategic focus (e.g., Berger and Ofek 1995; Comment and Jarrell 1995; Martin and Sayrak 2003). This "diversification discount" literature predicts that diversified insurers will be less efficient than focused insurers. We control for an insurers' diversification strategy through the use of Herfindahl indices, measured by net premiums written, that assess concentration based on product lines (*lbherfpw*) and participation across states (*pwherf*).

The regressions support the hypothesis that strategic focus is superior to diversification as a corporate strategy. The line of business and geographical Herfindahl indices are significant and positively related to cost, revenue, and profit efficiency, suggesting that diversified firms are less efficient than strategically focused firms. This implies that the benefits from risk diversification tend to be offset by the extra costs arising from management coordination, allocation of resources, and dealing with various regulations when operating in multiple lines and states. However, in terms of productivity improvement, diversified firms are able to surpass focused firms in reaping the benefits.

Distribution Systems: The relationship between distribution systems and efficiency in the insurance industry has been analyzed extensively in the literature (e.g., Kim et al. 1996; Regan and Tennyson 1996; Regan 1997). The traditional finding has been that firms relying on impendent agent channels tend to have higher costs than direct writers (Joskow 1973; Cummins and VanDerhei 1979; Barrese and Nelson 1992). However, Berger et al. (1997) provide evidence that, although independent distributors tend to be less cost efficient, the higher costs represent the provision of additional services that are valued by buyers such that independent and direct distributors are roughly equal in revenue and profit efficiency. Firms using the direct writing distribution systems are predicted to be more efficient to the extent that such firms can more easily recognize cost savings through automation and more fully realize the benefits of technology investments in customer databases and marketing than those insurers using less vertically integrated distribution systems (Carr et al. 1999). We control for the distribution system employed by including a dummy variable that equals one if the insurer uses a direct writing system (*direct*).

The two-way fixed effects regression results show that direct writers are more revenue and profit efficient than other types of channels but show no significant difference in cost efficiency. These findings are contrary to that of Berger et al. (1997) and calls further in-depth analysis of the issue. Our results could be attributable to our using a more up-to-date sample period and using DEA rather than stochastic frontier analysis.

Organizational and Ownership Form: Previous research has established that agency costs vary across different organizational forms (e.g., Fama and Jensen 1983; Mayers and Smith 1988). Stock insurance companies face inefficiency in the form of a misalignment of goals between management and ownership. However, the market for corporate control in stock firms mitigates managerial discretion distortions, forcing managers to maximize performance (Mayers and Smith 1981). Mayers and Smith (1988) suggest that mutual firms have far less pressure to maximize firm performance (at the expense of the policyholder), because the role of owner and policyholder functions are merged. The univariate analyses between efficiency and firm organizational form verify these predictions.

To control for organizational form, we include in the regression a variable identifying the organizational structure of the insurer (*mutual*). Concurring with the significant body of earlier research on the subject in the US market, we anticipate that a stock organizational form will display relatively higher levels of efficiency than its mutual counterpart, all other things equal. To our surprise, for the period studied, after controlling for firm fixed effects, mutual and stock insurers are found to have a similar capacity in achieving efficient operation, which somewhat explains the coexistence of the two types of insurers in the market.

We also control for the public trading status of insurers in the regression to examine the merits of having public ownership. The prediction of the effect of this variable on firm efficiency is again ambiguous, as argued in Sect. 6.3.5.2. Once controlling for other characteristics of firms, we find that public firms are no less efficient than private firms in all types of efficiency measures in the regressions. This is consistent with the findings in Xie (2010).

6.4 Conclusion and Discussion

This chapter examines efficiency and productivity for US P-L insurers during the period 1993–2011. The investigation is motivated by dynamic structural changes in the industry over the past two decades. Pure technical, scale, allocative, cost, and revenue efficiency, as well as profit efficiency, are estimated for the largest available sample of insurers using DEA. We investigate the organizational and ownership form, product range, and distribution channels in the market over time and provide analyses on the relationship between firm efficiency and firm characteristics. Panel data regression analysis is performed to explore the relationships in a multivariate regression framework. The study distinguishes itself from previous literature by using a longer time horizon of data and by providing profit efficiency– the ultimate goal of business operation – estimation of insurers.

On average, the industry operates with low cost and revenue efficiency, averaging 51 % and 44 %, respectively. The average profit inefficiency of the industry during the sample period is 0.63. These figures have improved slightly over time, especially during the recent years after the financial crisis. In addition, the efficiency scores in this industry are widely dispersed among firms, with a significant percentage of firms in the industry at the lower end, doing a poor job in minimizing costs and maximizing revenues. However, a good sign is that the industry is able to improve its operating efficiency over time, as reflected in the improvement of total factor productivity changes. In recent years, technical change (shift in production frontier) has been the primary driving force of the TFP improvement.

As most previous studies, medium sized firms are found to be more scale efficient. A significant percentage of firms in the industry operate with DRS, with an overwhelming majority of them being big firms in the last five asset deciles. Despite the benefits of scale economies for mergers and acquisitions among small to medium insurers, this study finds M&A deals among large insurers may be questionable in terms of improving scale efficiency. The finding provides additional support for challenging the expansion of insurers identified as G-SII (Global Systemically Important Insurers).

Further analyses on the factors affecting firm efficiency provide more insights regarding the relationship between insurer efficiency and its product, ownership form, distribution channels, and other business strategies. We have a stable percentage of mutual firms operating in the market, despite the fact that stock insurers are gaining slightly higher market shares (5 % from 1993 to 2011) over the last two decades. More insurers in the US P-L industry are found to become more focused on their core product lines and to employ multiple distribution channels to adapt to the rapid development of communication technologies to generate more premium inflows. A smaller percentage of insurers chose to stay publicly traded, which can possibly be attributed to the higher regulation and disclosure costs brought by the Sarbanes–Oxley Act of 2002 and the recent Dodd-Frank Act of 2010.

The conclusion one can draw about the determinants of a firm's efficiency is model dependent. Ignoring firm fixed effects in the regression can sometimes lead to different conclusions for some important variables. Given the facts that models considering firm fixed effects have better statistical properties, such models should be more reliable. In terms of profit maximization ability, the two-way fixed effects regressions indicate that personal lines insurers tend to be more efficient than commercial lines insurers. Higher diversification across either product lines or geographical areas is associated with lower efficiency, supporting the argument that strategic focus is a better strategy than diversification. Larger firms and firms with higher financial leverage and underwriting leverage tend to be more efficient, reflecting better capital management. Firms using direct writing distribution systems appear to be more efficient than those using agents and other distribution systems. The finding that ownership forms are not significantly related to firm efficiency can help explain the coexistence of various types of ownership in the market; however, the coexistence of the distribution channel puzzle still exists, which calls for more thorough study in the future.

References

Aly HY, Grabowski R, Pasurka C, Rangan N (1990) Technical, scale, and allocative efficiencies in US banking: an empirical investigation. Rev Econ Statist 72:211–218

Arrow K (1971) Essays in the theory of risk bearing. Markham, Chicago

Banker RD (1993) Maximum likelihood, consistency and data envelopment analysis: a statistical foundation. Manage Sci 39:1265–1273

Banker RD, Natarajan R (2008) Evaluating contextual variables affecting productivity using data envelopment analysis. Oper Res 56:48–58

Barrese J, Nelson JM (1992) Independent and exclusive agency insurers: a reexamination of the cost differential. J Risk Insur 59(3):375–397

Barros CP, Barroso N, Borges MR (2005) Evaluating the efficiency and productivity of insurance companies with a malmquist index: a case study for Portugal. Geneva Papers: Issues Practices 30:244–267

Berger AN, Humphrey DB (1992) Measurement and efficiency issues in commercial banking. In: Griliches Z (ed) Output measurement in the service sectors. University of Chicago Press, Chicago

Berger PG, Ofek E (1995) Diversification's effect on firm value. J Financ Econ 37(1):39–65

Berger AN, Cummins JD, Weiss MA (1997) The coexistence of multiple distribution systems for financial services: the case of property-liability insurance. J Bus 70:515–546

Biener C, Eling M (2011) The performance of microinsurance programs: a data envelopment analysis. J Risk Insur 78(1):83–115

Biener C, Eling M (2012) Organization and efficiency in the international insurance industry: a cross-frontier analysis. Eur J Oper Res 221(2):454–468

Brau JC, Fawcett SE (2006) Initial public offerings: an analysis of theory and practice. J Financ 61:399–435

Brockett PL, Cooper WW, Golden LL, Rousseau JJ, Wang Y (2004) Evaluating solvency versus efficiency performance and different forms of organization and marketing in US property—liability insurance companies. Eur J Oper Res 154(2):492–514

Brockett PL, Cooper WW, Golden LL, Rousseau JJ, Wang Y (2005) Financial intermediary versus production approach to efficiency of marketing, distribution systems, and organizational structure of insurance companies. J Risk Insur 72:393–412

Carr RM, David Cummins J, Regan L (1999) Efficiency and competitiveness in the US life insurance industry: corporate, product, and distribution strategies. In: David Cummins J, Santomero AM (eds) Changes in the life insurance industry: efficiency, technology and risk management. Kluwer, Boston

Casu B, Girardone C, Molyneux P (2004) Productivity change in European banking: a comparison of parametric and non-parametric approaches. J Bank Financ 28:2521–2540

Chen L-R, Lai GC, Wang JL (2011) Conversion and efficiency performance changes: evidence from the U.S. property-liability insurance industry. Geneva Risk Insur Rev 36:1–35

Comment R, Jarrell GA (1995) Corporate focus and stock returns. J Financ Econ 37:67–87

Cooper WW, Seiford LM, Tone K (2000) Data envelopment analysis: a comprehensive text with models, applications, references and DEA-solver software. Kluwer, Boston
Cooper WW, Seiford LM, Zhu J (2004) Handbook of data envelopment analysis. Kluwer, Boston
Cummins JD (1999) Efficiency in the US life insurance industry: are insurers minimizing costs and maximizing revenues? In: Cummins JD, Santomero AM (eds) Changes in the life insurance industry: efficiency, technology and risk management. Kluwer, Boston
Cummins JD, Danzon PM (1997) Price shocks and capital flows in liability insurance. J Financ Intermed 6:3–38
Cummins JD, Nini GP (2002) Optimal capital utilization by financial firms: evidence from the property-liability insurance industry. J Financ Serv Res 21:15–53
Cummins JD, Phillips RD (2005) Estimating the cost of equity capital for property-liability insurers. J Risk Ins 72(3):441–478
Cummins JD, Rubio-Misas M (2006) Deregulation, consolidation, and efficiency: evidence from the Spanish insurance industry. J Money Credit Bank 38:323–355
Cummins JD, Van Derhei J (1979) A note on the relative efficiency of property-liability insurance distribution systems. Bell J Econ 10(2):709–719
Cummins JD, Weiss MA (1993) Measuring cost efficiency in the property-liability insurance industry. J Bank Financ 17:463–481
Cummins JD, Weiss MA (2000) Analyzing firm performance in the insurance industry using frontier efficiency and productivity methods. In: Dionne G (ed) Handbook of insurance. Kluwer, Boston
Cummins JD, Weiss MA (2013) Analyzing firm performance in the insurance industry using frontier efficiency and productivity methods. In: Dionne G (ed) Handbook of insurance, 2nd edn. Kluwer, Boston
Cummins JD, Weiss MA (2014) Systemic risk and the U.S. insurance sector. J Risk Insur 81(3):489–527
Cummins JD, Xie X (2008) Mergers and acquisitions in the US property-liability insurance industry: productivity and efficiency effects. J Bank Financ 32(1):30–55
Cummins JD, Xie X (2009) Market values and efficiency in U.S. property-liability insurer acquisitions and divestitures. Manag Financ 35:128–155
Cummins JD, Xie X (2013) Efficiency, productivity, and scale economies in the U.S. property-liability insurance industry. J Prod Anal 39:141–164
Cummins JD, Zi H (1998) Comparison of frontier efficiency methods: an application to the US life insurance industry. J Prod Anal 10:131–152
Cummins JD, Tennyson S, Weiss MA (1999a) Consolidation and efficiency in the US life insurance industry. J Bank Financ 23:325–357
Cummins JD, Weiss MA, Zi H (1999b) Organizational form and efficiency: an analysis of stock and mutual property-liability insurers. Manage Sci 45:1254–1269
Cummins JD, Rubio-Misas M, Zi H (2004) The effect of organizational structure on efficiency: evidence from the Spanish insurance industry. J Bank Financ 28(12):3113–3150
Cummins JD, Weiss MA, Xie X, Zi H (2010) Economies of scope in financial services: a DEA efficiency analysis of the US insurance industry. J Bank Financ 34:1525–1539
Davutyan N, Klumpes PJM (2009) Consolidation and efficiency in the major European insurance markets: a non discretionary inputs approach. Working paper, Imperial College, London
Diacon SR, Starkey K, O'Brien C (2002) Size and efficiency in European long-term insurance companies: an international comparison. Geneva Papers Risk Insurance—Issues Practice 27(3):444–466
Eling M, Luhnen M (2010) Efficiency in the international insurance industry: a cross-country comparison. J Bank Financ 34:1497–1509
Erhemjamts O, Leverty JT (2010) The demise of the mutual organizational form: an investigation of the life insurance industry. J Money Credit Bank 42(6):1011–1036
Fama EF, Jensen MC (1983) Separation of ownership and control. J Law Econ 26:391–325

Färe R, Grosskopf S, Knox Lovell CA (1985) The measurement of efficiency of production. Kluwer-Nijhoff, Boston

Färe R, Grosskopf S, Norris M, Zhang Z (1994) Productivity growth, technical progress, and efficiency change in industrialized countries. Am Econ Rev 1994:66–83

Farrell MJ (1957) The measurement of productive efficiency. J R Statist Soc Ser A (General) 120:253–290

Fukuyama H, Weber WL (2001) Efficiency and productivity change of non-life insurance companies in Japan. Pacific Econ Rev 6:129–146

Gardner LA, Grace MF (1993) X-efficiency in the US life insurance industry. J Bank Financ 17:497–510

Grace MF, Leverty JT, Phillips RD, Shimpi P (2010) The value of investing in enterprise risk management. Working paper, Georgia State University

Grosskopf S (1996) Statistical inference and nonparametric efficiency: a selective survey. J Prod Anal 7:161–176

He E, Sommer DW, Xie X (2011) The impact of CEO turnover on property-liability insurer performance. J Risk Insur 78:583–608

Huang W, Eling M (2013) An efficiency comparison of the non-life insurance industry in the BRIC countries. Eur J Oper Res 226(3):577–591

Huang L-Y, Lai GC, McNamara M, Wang J (2011) Corporate governance and efficiency: evidence from U.S. property-liability insurance industry. J Risk Insur 78(3):519–550

Hughes JP, Mester LJ (1998) Bank capitalization and cost: evidence of scale economies in risk management and signaling. Rev Econ Statist 80:313–325

Ibbotson Associates (2012) Stocks, bonds, bills, and inflation: valuation edition 2012 yearbook. Morningstar, Chicago

Jeng V, Lai GC (2005) Ownership structure, agency costs, specialization, and efficiency: analysis of keiretsu and independent insurers in the Japanese nonlife insurance industry. J Risk Insur 72(1):105–158

Jeng V, Lai GC, McNamara MJ (2007) Efficiency and demutualization: evidence from the U.S. life insurance industry in the 1980s and 1990s. J Risk Insur 74(3):683–711

Joskow PL (1973) Cartels, competition and regulation in the property-liability insurance industry. Bell J Econ Manage Sci 4(2):375–427

Kim W-J, Mayers D, Smith CW Jr (1996) On the choice of insurance distribution systems. J Risk Insur 63(2):207–227

Kittelsen S (1995) Monte Carlo simulations of DEA efficiency measures and hypothesis test. MEMORANDUM, No 09/99, Department of Economics, University of Oslo

Kneip A, Park BU, Simar L (1998) A note on the convergence of nonparametric DEA estimators for production efficiency scores. Econ Theory 14:783–793

Lemaire J (1985) Automobile insurance: actuarial models. Kluwer-Nijhoff, Boston

Lemaire J (1995) Bonus-Malus system in automobile insurance. Kluwer-Nijhoff, Boston

Leverty JT, Grace MF (2010) The robustness of output measures in property-liability insurance efficiency studies. J Bank Financ 34(7):1510–1524

Leverty JT, Grace MF (2012) Dupes or incompetents? An examination of management's impact on firm distress. J Risk Insur 79(3):751–783

Ma Y-L, Pope N, Xie X (2013a) Contingent commissions, insurance intermediaries, and insurer performance. Risk Manage Insur Rev 17(1):61–81

Ma Y-L, Pope N, Xie X (2013b) Insurer performance and intermediary remuneration: the impact of abandonment of contingent commissions. Geneva Papers Risk Insurance—Issues Practice 39(2):373–388

Mahlberg B, Url T (2003) Effects of the single market on the Austrian insurance industry. Empir Econ 28:813–838

Mahlberg B, Url T (2010) Single market effects on productivity in the German insurance industry. J Bank Financ 34:1540–1548

Martin JD, Sayrak A (2003) Corporate diversification and shareholder value: a survey of recent literature. J Corporate Fin 9(1):37–57
Mayers D, Smith CW (1981) Contractual provisions, organizational structure, and conflict control in insurance markets. J Bus 54(3):407–434
Mayers D, Smith CW Jr (1987) Corporate insurance and the underinvestment problem. J Risk Insur 54(1):45–54
Mayers D, Smith CW (1988) Ownership structure across lines of property-casualty insurance. J Law Econ 31(2):351–378
McAllister PH, McManus D (1993) Resolving the scale efficiency puzzle in banking. J Bank Financ 17:389–405
Ray S, Desli E (1997) Productivity growth, technical progress, and efficiency change in industrialized countries: comment. Am Econ Rev 87:1033–1039
Regan L (1997) Vertical integration in the property-liability insurance industry: a transaction cost approach. J Risk Insur 64(1):41–62
Regan L, Tennyson S (1996) Agent discretion and the choice of insurance marketing system. J Law Econ 39(2):637–666
Shephard RW (1970) Theory of cost and production functions. Princeton University Press, Princeton
Simar L, Wilson PW (1998) Productivity growth in industrialized countries. Discussion paper 9810, Institut de Statistique, Université Catholique de Louvain, Belgium
Taylor G (2000) Loss reserving: an actuarial perspective. Kluwer, Boston
Wang JL, Jeng V, Peng JL (2007) The impact of corporate governance structure on the efficiency performance of insurance companies in Taiwan. Geneva Papers: Issues Practices 32:264–282
Xie X (2010) Are publicly held firms less efficient: evidence from the US property-liability insurance industry. J Bank Financ 34:1549–1563
Xie X, Lu W, Reising J, Stohs MH (2011) Demutualisation, control and efficiency in the U.S. life insurance industry. Geneva Papers: Issues Practices 36:197–225
Yuengert AM (1993) The measurement of efficiency in life insurance: estimates of a mixed normal-gamma error model. J Bank Financ 17:483–496

Chapter 7
Mutual Fund Industry Performance: A Network Data Envelopment Analysis Approach

I.M. Premachandra, Joe Zhu, John Watson, and Don U.A. Galagedera

Abstract The objective of this chapter is twofold. First, we present a comprehensive review of the DEA literature that has evaluated mutual fund performance. Second, we present a two-stage DEA model that decomposes the overall efficiency of a decision-making unit into two components and demonstrate its applicability by assessing the relative performance of 66 large mutual fund families in the US over the period 1993–2008. By decomposing the overall efficiency into operational management efficiency and portfolio management efficiency components, we reveal the best performers, the families that deteriorated in performance, and those that improved in their performance over the sample period. We also make frontier projections for poorly performing mutual fund families and highlight how the portfolio managers have managed their funds relative to the others during financial crisis periods.

Part of this chapter is based upon materials from Premachandra, I.M., Zhu, J., Watson, J., and Galagedera, D.U.A., 2012. Best-performing US mutual fund families from 1993 to 2008: Evidence from a novel two-stage DEA model for efficiency decomposition, *Journal of Banking and Finance*, 36, 3302–3317.

I.M. Premachandra
Department of Accountancy and Finance, School of Business University of Otago, Dunedin, New Zealand
e-mail: i.premachandra@otago.ac.nz

J. Zhu
International Center for Auditing and Evaluation, Nanjing Audit University, Nanjing, P.R., China

School of Business, Worcester Polytechnic Institute, Worcester, MA 01609-2247, USA
e-mail: jzhu@wpi.edu

J. Watson
Department of Banking and Finance, Monash Business School, Monash University, Melbourne, VIC, Australia
e-mail: jr.watson@monash.edu

D.U.A. Galagedera (✉)
Department of Econometrics and Business Statistics, Monash Business School, Monash University, Melbourne, VIC, Australia
e-mail: tissa.galagedera@monash.edu

J. Zhu (ed.), *Data Envelopment Analysis*, International Series in Operations Research & Management Science 238, DOI 10.1007/978-1-4899-7684-0_7

Keywords Operational management efficiency • Portfolio management efficiency • Data envelopment analysis • Input–output models • Mutual fund families • Performance

7.1 Introduction

The mutual fund industry in the US is by far the largest such industry in the world, managing US$14.3 trillion in assets by the end of the calendar year 2012. Research on performance at the mutual fund family level is limited (Tower and Zheng 2008; Elton et al. 2007), possibly due to the complex nature of the analysis involved. Despite the limited existing research to date, an understanding of performance (absolute and relative) at the fund family level is important as investors tend to invest in funds within the same mutual fund family rather than across a number of families. The reasons for investing within one mutual fund family include convenience in searching for investment opportunities and recordkeeping (Kempf and Ruenzi 2008) and flexibility of switching funds without additional sales charges and restrictions imposed by the fund family (Elton et al. 2006, 2007).

Mutual fund performance receives substantial coverage in much of the US financial press due to the rapid growth of the mutual fund industry as well as the vital role it plays in the financial market. Investors and media commentators are keen to acquire an enhanced understanding of operational aspects at both the fund level and the fund family level given the recent turmoil experienced in the US financial market. This chapter gives an overview of the US mutual fund industry for open funds and respond to the line of criticism faced by the standard DEA-models by using a two-stage network DEA model that decomposes the overall efficiency of a fund family into two components; an operational management efficiency and portfolio management efficiency and thereby making a contribution to the mutual fund performance appraisal literature and mutual fund industry at large. We demonstrate the application of the proposed DEA model by examining the relative performance of 66 large mutual fund families in the US over the period 1993–2008.

We conceptualise the activities of mutual fund management as a two-stage process as follows. In the first stage, we focus on the operational management aspect and investigate how efficiently the managers at the fund family level make use of inputs such as marketing and distribution expenses and management fees in producing the output, which is the net asset value. In the second stage, the focus is on the portfolio management aspect where we determine how efficiently the fund managers make use of inputs such as fund size, standard deviation of the returns, turnover ratio, expense ratio and net asset value in producing the output, which is fund family average return. Brown et al. (2001) point out that even though relative performance appears to be the overriding concern of fund managers as well as their clients, considerably less attention is directed towards the equally important question of relative performance appraisal of portfolios.

We treat net asset value (NAV) which is considered as the output variable at the first stage as an input variable in the second stage; that is, net asset value is modeled as an intermediate variable that links stage 1 with stage 2. Holod and Lewis (2011) treat deposits in the same way in the two-stage network DEA model they use in assessing bank performance. Our modelling framework aligns with the network structure of Färe and Whittaker (1995). Although we consider only one output from the first stage and one output from the second stage in this particular application, the DEA model that we use here allows multiple inputs, outputs and intermediate measures (Premachandra et al. 2012).

Our model splits the overall process of a DMU into two stages and assesses the efficiencies of both stages simultaneously. Our two-stage network DEA model not only assesses the overall performance of the DMUs, but also decomposes the overall efficiency into two components associated with the performance in the two stages. Such a decomposition of overall efficiency is not possible in the previous network approach by Färe and Whittaker (1995). Furthermore, our modelling framework allows assessments under the variable returns to scale (VRS) as well as constant returns to scale (CRS) assumptions and as such it is not restrictive in terms of orientation as in Kao and Hwang's (2008) two-stage model, which is valid only under the CRS assumption. Usually, in the two-stage DEA models, the intermediate variables that link stage 1 with stage 2 become the inputs of stage 2. Our model allows new variables as inputs in the second stage in addition to the intermediate variables. Interested reader is referred to Cook and Zhu (2014) for recent developments in network DEA modeling techniques.

The chapter is structured as follows: Sect. 7.2 provides an overview of the US mutual fund industry, Sect. 7.3 examines the literature on mutual fund performance appraisal, Sect. 7.4 formulates the two-stage DEA model and in Sect. 7.5 the data used in the application are presented. Section 7.6 analyses the fund family performance and Sect. 7.7 presents concluding remarks.

7.2 Background to US Mutual Fund Industry

According to the Investment Company Institute (ICI 2013), the mutual fund industry (MFI) in the United States is by far the largest such industry in the world (see Fig. 7.1), managing \$13.1 trillion in assets as at the end of 2012 which accounts for 48.9 % of the \$26.8 trillion worldwide value of assets under management in the industry. There has been a significant growth in US mutual fund industry over the 10 years from 2003 having almost doubled the total market value of assets under management to \$7.4 trillion. The total value of funds under management in the US industry has rebounded since the onset of the global financial crisis; increasing by 25 % since 2006. Measuring the growth of the MFI is much more complex than simply looking at the growth in dollar value of assets under management. Other dynamic measures such as net flow of funds into mutual funds (MFs) also matter.

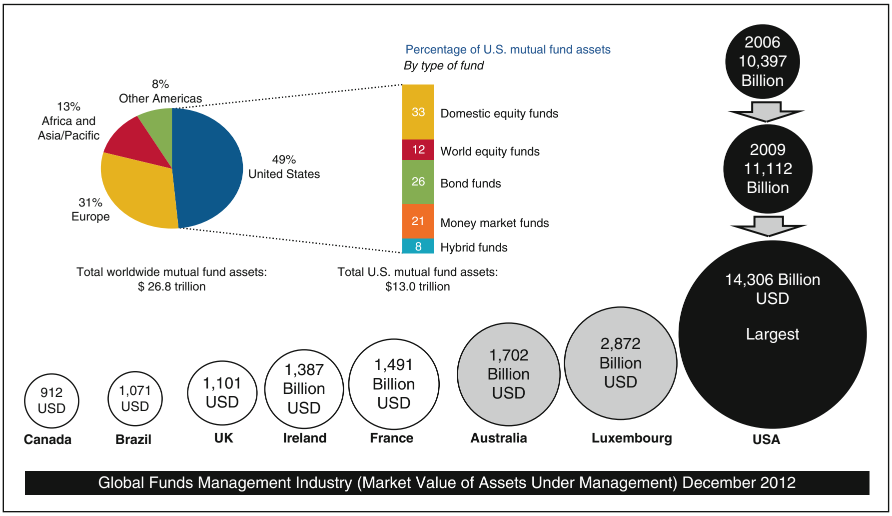


Fig. 7.1 Global significance of United States Mutual Fund Asset Pool (December 2012). Source of data: Investment Company Fact Book 2010, Worldwide Total Net Assets of Mutual Funds. This figure reports the size of investment fund industries around the world. All dollar values are represented in billions of US dollars at the end of the 2012 calendar year

At year-end 2012 (ICI 2013), the number of fund products constituting the US MFI was approximately 7596 sponsored by more than 700 fund families. Nevertheless, since the dawn of the new millennium the percentage of industry assets invested in larger fund complexes has increased. The share of the assets managed by the largest ten US fund families in 2012 was 53 %, up from the 44 % in 2000. Long run competitive dynamics have prevented any single fund or family of funds from dominating the market. For example, out of the largest 25 fund complexes in 1995, only 15 remained at the top level in 2012. The composition of the assets held in the top 25 fund complexes has changed significantly with a relative reduction in domestic equity holdings and an increase in money market funds. Nevertheless, this could be representative of the financial situation that prevail post 2008 meltdown. The Herfindahl Hirschman Index for the US MFI is 465 (ICI 2013, p. 25) which is well below the 1000 that is considered as the cutoff for a concentrated industry. To this end, it is deemed that the US MFI still offers to investors products that vary significantly in size, number of investment classes, investment horizon, and management style.

7.3 Prior Research on Performance Appraisal of Mutual Funds and Mutual Fund Families

One of the major motivations for mutual fund managers, whether at the fund family level or at the fund product level, is to maintain high standard of performance compared to their peers so that in the event of a temporary setback they are able to manage possible cash outflows and potential job losses better. This is a vital challenge for fund family managers in the US due to the increasing competition within the mutual fund industry. Maintaining high standard of performance is consistent with fund family managers minimizing controllable efforts (inputs) to achieve the highest possible level of return (outputs) defined by a production frontier (Charnes et al. 1978). This phenomenon which is consistent with the economic theory on optimization provides a strong motivation for adopting the production frontier concept in performance appraisal of mutual funds.

Fund family managers aim for their products to lie on the outer extremities of the production frontier so that their funds are more efficient than the other funds of comparable type. However, in reality, they may fall short due to reasons within and sometimes beyond their control. It is this notion of a shortfall of performance of some mutual funds relative to other funds in the sample that aligns with the concept of production inefficiency which is a measurable quantity.

In the past 25 years, innovative approaches have been introduced to measure mutual fund performance at both the individual fund product level (Murthi et al. 1997; McMullen and Strong 1998) and more recently at the fund family level (Premachandra et al. 2012). In general, the findings support the assertion that fund managers should be concerned about inefficiencies not only in managing funds but also in their operations. Table 7.1 presents a summary of the used DEA model, the input and output variables and the key findings of some of the significant studies conducted on mutual fund performance appraisal since the pioneering work in the investment funds area by Murthi et al. (1997).

DEA has the following unique features. First, DEA does not require a priori assumption on the relation between inputs and outputs. It can handle multiple performance measures classified as inputs and outputs in a single mathematical model without the need for trade-off between the inputs and outputs associated with performance. The literature has shown DEA to be a valuable instrument for performance evaluation and benchmarking (Zhu 2002; Cooper et al. 2004). Second, DEA examines each DMU independently by generating individual performance (efficiency) scores that are relative to the DMUs in the entire sample under investigation. Misspecification, a recurring problem in regression analysis, is not a concern with DEA models since DEA creates the best practice frontier based on comparison of the peers in the sample. Third, it has been documented that DEA can assist with the study of a frontier shift over a time horizon, using for example, the DEA-based Malmquist index of Fare et al. (1997). This allows exploration of the dynamic change of fund family failure or success over time. The fourth advantage is that DEA does not need a large sample size (usually required by statistical and econometric approaches) for the evaluation of mutual funds or mutual fund families. The need for large sample sizes is a significant drawback when investment decisions have to be made using smaller samples. DEA can bypass such practical difficulties (see Premachandra et al. 2009).

Selection of performance measures or input and output factors is an important issue in DEA application (see Cook et al. 2014). For risk-averse investors, the capital market theory dictates that the higher the risk that you take in the investment, the

Table 7.1 Summary of relevant literature

Year	Authors	Publication	Sample period (sample size) and country/region	Inputs (I) and outputs (O) used in the DEA analysis	DEA model used and findings
1997	Murthi, B.P.S., Choi, Y.K., and Desai, P.	*European Journal of Operational Research*	1993 (2083) USA	Standard deviation (I), expense ratio (I), total load (I), turnover (I) and annual return (O)	DEA portfolio efficiency index constructed using the CCR-O model (assume CRS); Propose a new measure of performance to overcome shortcomings of Jensen's alpha and Sharpe index to address the limitations of these metrics. (Find that mutual funds are all approximately mean-variance efficient)
1998	McMullen, P.R., and Strong, R.A.	*The Journal of Business and Economic Studies*	1997 (135) USA	3-Year standard deviation (I), sales charge (I), minimum investment (I), and expense ratio (I), 1 year return (O), 3 year return (O) and 5 year return (O)	CCR-O with restricted weights (assume CRS); Purpose of the study is to demonstrate the DEA methodology rather than prescribe an alternative to mean-variance optimisation. (Find that DEA can assist in selecting mutual funds for investors with multifactor utility functions)
1998	Premachandra, I.M., Powell, J.G., and Jing Shi	*Omega, International Journal of Management Science*	1975–1992 (16) New Zealand	Asset allocation in risky holdings (I), Asset allocation in risk free asset (I), total market value at end of time period (O)	Stochastic DEA model; The paper proposes a spread-sheet based numerical model formulated to alleviate short-term performance concerns within the New Zealand portfolio management sector. (Find that while DEA is useful in the

					selection of portfolios, SDEA approach is more appropriate given the chance element in short term portfolio performance)
2000	Powers, J.	*Journal of Business and Management*	1998 (185) USA	P/E ratio (I), 5 year systematic risk (I), 5 year standard deviation (I)/ EPS (O), 1 year return (O),3 year return (O),5 year return (O), and 10-year return (O)	Use both the CCR-O model (assume CRS) and BCC-O model (assume VRS); Shows how DEA can be used to assist with the multi-criteria problem in stock selection. (Find that 7.5 % of large cap stocks are efficient and an additional 2 % are desirable)
2001	Basso, A., and Funari, S.	*European Journal of Operational Research*	1997–1999 (47) Italy	Subscription costs (I), redemption fees (I), standard deviation (I), square root of half variance (I),beta (I), 1 year return (O) and 3 year return (O)	Generalised DEA portfolio efficiency index constructed using the CCR-O model (assume CRS); Present a model (that defines mutual fund performance as an index) which can be used to evaluate the performance of mutual funds. (For each mutual fund, the procedure identifies a composite portfolio which may be considered as a benchmark)
2001	Choi, Y.K., and Murthi, B.P.S.	*Journal of Business Finance and Accounting*	1993 (731) USA	Expense ratio (I), Loads (I), 3 year standard deviation (I), turnover (I), Net asset value (I), and 3 year return (O)	BCC-O model (assume VRS) and CCR-O model (assume CRS); Present an alternative mutual fund performance evaluation measure that does not require any functional form for

(continued)

Table 7.1 (continued)

Year	Authors	Publication	Sample period (sample size) and country/region	Inputs (I) and outputs (O) used in the DEA analysis	DEA model used and findings
					the return/risk or return/cost relationship. (The index presented is a variant of the Sharpe index as it measures not only the performance per unit of risk but also the performance per unit of cost)
2002	Galagedera, D.U.A., and Silvapulle, P.	*Managerial Finance*	1995–1999 (257) Australia	Standard deviations of the 1-, 2-, 3- and 5-year gross performance (I); sales charges (I); Management Expense Ratio (I); minimum initial investment (I). Short-term performance in the last 12 months (O); medium-term performances include 2- and 3 year gross performances (O). The ex post 5-year gross performance reflects output in long term (O)	BCC-I model (assume VRS); Suggest that DEA techniques can overcome some of the problems of the capital asset pricing model. Finds a positive association between ratings and DEA efficiency scores
2003	Basso, A., and Funari, S.	*The Journal of the Operational Research Society*	Empirical application (50) funds randomly generated (30 non and 20 ethical funds)	Subscription costs (I); redemption costs (I); standard deviation (I); beta (I); expected return (O); and ethical indicator (O)	DEA portfolio efficiency index constructed using the CCR-O model (assume CRS); Propose three models that tackle the problem of negative average rates of return
					Evaluates the performance of ethical mutual funds

2003	Haslem, J.A., and Scheraga, C.A.	*The Journal of Investing*	1999 (80) USA	Cash [%] (I); Expense ratio (I); stocks [%] (I); P/E ratio (I); P/B ratio (I); fund total assets (I); and Sharpe Index (O)	CCR-I model (assume CRS); Find that the investment style of large cap DEA-efficient funds are predominantly value funds rather than growth funds
2003	Sengupta, J.K.	*Applied Financial Economics*	1988–1999 (60) USA	Load (I); expense ratio (I); turnover (I); standard deviation of return (I); and the return covariance with the S&P 500 (I); mean return; (O) and skewness (O)	BCC-I model (assume VRS); Finds that mean variance efficiency hypothesis holds for approximately 75 % of the sample. Also finds that, among the efficient funds the technology and communication fund has second degree stochastic dominance over the growth fund and exhibits a positive probability of beating the market in terms of probability dominance
2004	Anderson, R.I., Brockman, C.R., Giannikos, C., and McLeod, R.W.	*International Journal of Business and Economics*	1997–2001 (348) USA	Front Load (I); Deferred Load (I); 12b-1 fees (I); Other expenses (I); Standard deviation (I); and annual return (O)	CCR-O model (assume CRS); Determine the sources of inefficiencies by examining the mean inefficiencies of the input and output values. Report that the consistently significant inefficiencies were found only with the RMF's loads and 12b-1 fees
2004	Chang, K.P.	*Computers and Operations Research*	1992–1996 (701) USA	1 and 5 years Beta (I); 1 and 5 years standard deviation (I); assets in millions (I); and 1 and 5 years return (O)	Non-standard DEA model. Adopts minimum convex input requirement set approach; Find that maximum capital gain and growth funds have performed

(continued)

Table 7.1 (continued)

Year	Authors	Publication	Sample period (sample size) and country/region	Inputs (I) and outputs (O) used in the DEA analysis	DEA model used and findings
					worse than growth and income funds; actively managed funds underperform passive investment strategy; low risk funds outperform high risk funds and no load funds outperform load funds. Also finds that funds with low beta and small assets under management have operated more efficiently
2005	Basso, A., and Funari, S.	*Central European Journal of Operations Research*	1997–2001 (50) Italy	Standard deviation (I); beta coefficient (I); subscription costs (I); redemption costs(I); Sharpe measure (O); reward to half variance (O); Jensen alpha (O): and Treynor ratio (O)	A generalised DEA performance indicator; Propose a model that can be used to define relative performance of mutual funds which takes into account all different aspects considered in traditional performance metrics
2005	Gregoriou, G.N., Sedzro, K., and Zhu, J.	*European Journal of Operational Research*	1997–2001 (8) USA	lower mean monthly semi-skewness (I); lower mean monthly semi-variance;(I) mean monthly lower return (I); upper mean monthly semi-skewness (O); upper mean monthly semi-variance (O) and mean monthly upper return (O)	Super-efficiency model (Andersen and Petersen 1993); Suggest that DEA may be used as a complimentary technique in the selection of efficient hedge funds and funds of hedge funds. DEA can shed light and further validate hedge fund selection under other methodologies

2006	Haslem, J.A., and Scheraga, C.A.	*The Journal of Investing*	2001 (58) USA	Cash % (I); expense ratio (I); stocks (I); P/E Ratio (I); P/B ratio (I); number of securities held (I); portfolio turnover (I); and total assets (O)	CCR-I model (assume VRS); Find that mutual funds that are managerially inefficient tend to have the largest values for the seven investment style variables. Conclude that the growth style of portfolio management is more prone to being managed inefficiently
2006	Chen, Z., and Lin, R.	*Operations Research Spectrum*	1999–2002 (33) China	Standard deviation (I); Beta (I); square root of the lower semi-variance (I); value at risk (I); conditional value at risk (I); average transaction cost (I); return (O); and Jensen alpha (O)	Use both the CCR-I model (assume CRS) and BCC_I model (assume VRS); Find that VaR and CVaR, especially their combinations with traditional risk measures, are helpful for describing return distribution properties and identifying optimal fund characteristics such as the asset allocation structure. The authors infer that inclusion of VaR and CVaR allow for better evaluation of overall performance of mutual funds
2006	Daraio, C., and Simar, L.	*European Journal of Operational Research*	2001–2002 (5851) USA	Standard deviation (I); Expense Ratio (I); Loads (I); Turnover Ratio (I); and Return (O)	A robust nonparametric approach compared with CCR-I (assuming VRS); Find that most US mutual funds did not exploit the economies of scale deriving from portfolio management and shareholder services to a larger number of

(continued)

Table 7.1 (continued)

Year	Authors	Publication	Sample period (sample size) and country/region	Inputs (I) and outputs (O) used in the DEA analysis	DEA model used and findings
					securities/customers. Conclude that US funds learnt to deal with an increased number of securities/customers which offset the operational economies of scale. Claim that DEA is a powerful tool that has the ability to describe the effects of market risk such as the fallout from September 11 attacks
2006	Eling, M.	*Financial Markets and Portfolio Management*	1996–2005 (30) USA	Std deviation (I); maximum drawdown (I), average drawdown (I), std dev of drawdown (I), value at risk (I), conditional value at risk (I) and higher partial moment 1-3, average return (O), skewness (O) and minimum return (O)	Use the CCR-O model (assume CRS) the BCC-O model (assume VRS) and the super efficiency model (Andersen and Petersen 1993); Provides criteria for selecting inputs and outputs
2007	Hsu, C.L., and Lin, J.R.	*The Service Industries Journal*	1999–2003 (192) Taiwan	Mean total net assets (I); Mean Management fee (I); Mean load fee (I); Mean turnover ratio (I); standard deviation (I); and Total Return (O)	CCR-O model (assume CRS); Identify a significant 'hot hands' effect in Taiwan's domestic equity fund market. Therefore suggest that investors can benefit from chasing past winners and from avoiding past losers. Conclude that the difference in performance persistence between the above two measures is driven by the DEA

					taking both transaction costs and risk into consideration. Suggest that transaction costs play an important role in determining the performance of mutual funds
2008	Hu, J.L., and Chang, T.P.	*Applied Financial Economics Letters*	2005–2006 (156) USA	Standard deviation (I); expense ratio (I); and total return (O)	CCR-O model (assume CRS); Find that a fund's performance significantly increases with manager's tenure and education while it decreases with the number of funds managed. Also report that the relationship between the pure efficiency score and original score is positive
2008	Lozano, S., and Gutierrez, E.	*European Journal of Operational Research*	2002–2005 (108) Spain	Standard deviations (I); and Return (O)	Formulated return–risk DEA models and return–safety DEA models constructed using BCC-O (assume VRS); The DEA models proposed in this paper take into account the conflicting return and risk (or safety) aspects of performance assessment in a manner that does not neglect diversification effects. The proposed DEA approach can incorporate other factors such as entry fees and sales charges

(continued)

Table 7.1 (continued)

Year	Authors	Publication	Sample period (sample size) and country/region	Inputs (I) and outputs (O) used in the DEA analysis	DEA model used and findings
2011	Alexakis, P., and Tsolas, I.	*Multinational Finance Journal*	2001–2004 (55) Greece	Standard deviation (I); beta coefficient (I); assets and sales commissions or charges (I); and annualized daily arithmetic returns (O)	Use both the CCR-I model (assume CRS) and BCC-I model (assume VRS); Finds that the mean-variance efficiency hypothesis holds for the inefficient funds implying that these mutual funds had the highest expected return at their given level of risk
2011	Chen, Y.C., Chiu, Y.H., and Li, M.C.	*South Africa Journal of Economics*	2007 (278) Taiwan	Monthly purchasing turnover rate (I); Direct transaction cost rate(I); Selling expense rate(I); Monthly standard deviation (I); Treynor Index(O); Sharpe Index (O); Jensen Index(O); Monthly rate of return (O)	Use BCC-O model (assuming VRS); Findings show that (i) the BCC model and the system BCC model estimate significantly different efficiency scores (ii) under the system BCC model, balanced funds have larger average efficiency scores than stock funds (iii) there are more efficient funds under the system BCC model than under the BCC model and (iv) the number of funds with the same reference set is less than half of the total number of funds under both models
2011	Watson, J., Premachandra, I.M., and Wickramanayake, J.	*Managerial Finance*	1990–2005 (22) Australia	Total risk (I); beta (I); information ratio (I); cost (I); and stochastic monthly rate of return (O)	Stochastic DEA model; Find that Morningstar ratings in Australia provide investors with useful information.

					Advantages of the SDEA spread-sheet model include (i) easy replication as the model is created in Excel (ii) allows the user to make use of additional @RISK probability functions to model complicated relationships between input and output variables (iii) the output produced by the proposed method contains valuable statistical information about the random properties of the efficiency score of the DMU and (iv) the efficiency score distribution information can be used to compare DMU in alternative ways that suit individual user preferences
2012	Pendaraki, K.	*Journal of Applied Finance and Banking*	2007–2010 (43) Greece	Standard deviation (I); fund size (I); skewness (I); kurtosis (I); and Cumulative return (O)	BCC-I model (assume VRS); Find that higher moments provide a better measure of performance. Report that funds with high sensitivity to negative market conditions such as high kurtosis and skewness are more likely to generate lower efficiency scores
2012	Premachandra, I.M., Zhu, J., Watson, J., and Galagedera, D.U.A.	*Journal of Banking and Finance*	1990–2008 (66) USA	Management fees (I-1); Marketing and Distribution fees (I-1); NAV(O-1 and I-2) fund size (I-2); Net expense ratio (I-2);	A two stage DEA model (assume VRS); The 2-stage DEA model decomposes efficiency into operational and

(continued)

Table 7.1 (continued)

Year	Authors	Publication	Sample period (sample size) and country/region	Inputs (I) and outputs (O) used in the DEA analysis	DEA model used and findings
				turnover (I-2); standard deviation (I-2) and Average return (O-2)	portfolio efficiencies. Find that mutual fund families with good portfolio management did better during financial crisis periods
2012	Zhao, X and Yue, W.	*Procedia Computer Science*	2004–2008 (32) China	Subsystem of investment and research, weighted VAR during term 1 (I), weighted VAR during term 2 (I), the reverse of fund managers average tenure (I). Number of funds (O), Number of types (O), Product innovation speed (O), weighted return during term 1 (O), weighted return during term 2 (O), scale growth (O), average initial subscription scale (O), information service quality (O), total shares (O)	Formulate a Multi-subsystem Fuzzy DEA model. CCR-O (assume CRS) and BCC-O (assume VRS). Identify how close are the mutual fund management companies to the best practice frontier. Find that those companies that display relatively high managerial skills for the most part differ a lot in terms of marketing and service
2012	Rubio, J., Hassan, M., and Merdad, H.	*Accounting Research Journal*	2003–2010 (22,545); Islamic Funds (95), American funds (20,946), international funds (1504)	standard deviation (I), the lower partial momentums (I), and maximum drawdown period (I). Expected returns (O), the upper partial momentums (O), and the maximum period of consecutive gain (O)	BCC-I (assume VRS); There is strong evidence suggesting that Islamic funds are highly efficient and that they outperform their international counterparts

2014	Matallin, C., Soler, A., and Tortosa-Ausina, E.,	*Omega*	2001–2011 (1450) USA	Standard deviation (I) kurtosis (I), expense ratio (I), beta (I). daily mean return (gross return, y1) over the sample period as (O), skewness (O)	Non-convex counterpart of DEA (FDH) and order-m and order-α partial frontiers. BCC-O (assume VRS)
					Propose a method for testing the performance of DEA and FDH (Free Disposal Hull) methods in fund selection

This table gives information on studies of mutual fund performance assessed under the DEA framework. The table reports the names of the authors, place and date of publication, together with the sample period, sample size, the inputs and outputs used in the analysis, the type of model adopted and a summary of the main findings greater is the return. This implies a functional relationship between risk and return. Hence, in principle, it follows that risk measures may be considered as inputs in the DEA model and return measures as outputs. McMullen and Strong (1998) document the relation between risk and return and highlight that investors are concerned about risk and return over various time horizons as that allows investors to obtain greater information about a fund than simply looking at performance over a single time period. In addition to the risk–return trade-off, Murthi et al. (1997) report that investors are equally concerned about transaction costs such as subscription and redemption fees. Basso and Funari (2001, 2005) document that some investors also consider ethical criteria in their decisions. Thus, there is no consensus among researchers as to what input and output variables should be included in a DEA model when investigating the relative performance of mutual fund products.

In mutual fund performance appraisal, some of the input output factors considered in the DEA model such as the annual average return of a fund may take negative values. This problem can easily be resolved by translating such variables into positive values by adding a constant and then using an appropriate translation invariant DEA model. For example, the input-oriented BCC model (BCC-I) is translation invariant with respect to outputs, but not inputs. Similarly, the output oriented BCC model (BCC-O) is invariant under the translation of inputs, but not outputs. The additive DEA model is translation invariant in both inputs and outputs (See for example Cooper et al. (2006) for details). Table 7.1 lists various DEA models that have been used for mutual fund performance appraisal in the past. The standard DEA models do not account for the activities involved in transforming inputs into outputs and instead consider the DMU operation as a black box. In our case, we look inside this black box and consider the process of overall management of mutual fund families (the DMUs of our empirical application) as a combination of two sub processes namely; operational management and portfolio management.

7.4 Development of the Two-Stage DEA Model

Cook et al. (2010) document that in many instances, the underlying process of generating outputs from inputs may have a two-stage network structure with intermediate measures where outputs from the first stage become the inputs to the second stage. Chilingerian and Sherman (2004) describe such a two-stage process used in measuring physician care. Their first stage is a manager-controlled process and the second stage is a physician-controlled process. In their model, the output of the first-stage is considered as input to the second stage. The factors that link the two stages are called intermediate measures. Kao and Hwang (2008) consider the process of Taiwanese non-life-insurance companies as a two-stage process of premium acquisition and profit generation. In our application, we assume that the activities of mutual fund families can be viewed as a two-stage process where stage 1 represents the operational management process and stage 2 represents the portfolio management process. In the current application, the overall efficiency of a mutual fund

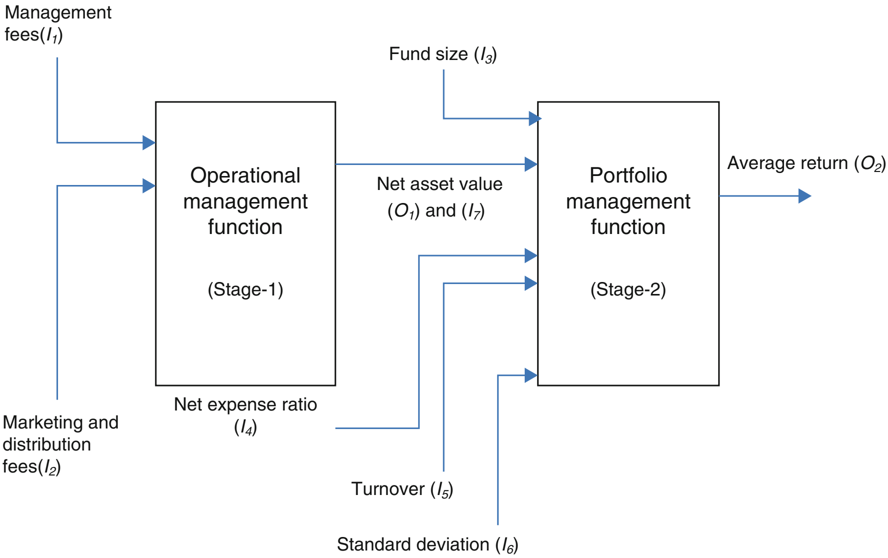


Fig. 7.2 The proposed two-stage DEA model for evaluating the efficiency of mutual fund families. At stage 1, the operational management efficiency will be estimated, and at stage 2 the portfolio management efficiency will be estimated. The overall efficiency of the fund family is decomposed into the operational management efficiency (stage 1) and the portfolio management efficiency (stage 2). Variables I_1 and I_2 are the input variables and O_1 is the output variable at stage 1 and I_3, I_4, I_5, I_6 and I_7 are the input variables and O_2 is the output variable at stage 2. Net asset value is an intermediate variable and therefore I_7 is the expected value of O_1 estimated in stage 1

family is conceptualized as made up of two components; operational management efficiency (hereinafter referred to as operational efficiency) and portfolio management efficiency (hereinafter referred to as portfolio efficiency). A schematic diagram of the mutual fund family management process is given in Fig. 7.2. In stage 1, the fund family management makes an attempt to attract funds from the investors and therefore outgoings such as management fees (I_1) and marketing and distribution expenses (I_2) that contribute directly towards generating funds are considered as the input variables. In stage 1 of Fig. 7.2, we consider the net asset value labeled O_1 as the output variable. Hence, a mutual fund family that produces the highest net asset value with the least amount of management fees and marketing and distribution expenses is considered to be operationally more efficient than the other families in the sample. Stage 2 is the portfolio management stage. Here we treat net asset value (O_1), fund size (I_3), net expense ratio (I_4), turnover ratio (I_5) and standard deviation of the returns of the family portfolio over the last 3 years (I_6) as the input variables and mean return of the family portfolio (O_2) as the output variable. Since net asset value (O_1), which is an output variable of stage 1 is also an input variable of stage 2 (I_7), it becomes an intermediate variable. I_7 is not observable; it is obtained by adjusting O_1 which is observed. In stage 2, a fund family that produces the highest average family portfolio return with the least amount of net asset value, fund size, net expense ratio, turnover ratio, and standard deviation is deemed more efficient compared to the other families in the sample.

A common approach to solving two-stage network problems illustrated in Fig. 7.2 is to assume that the two stages operate independently and apply a standard DEA model separately in each stage. Various problems could arise due to this approach. For example, in stage 1, a fund may attempt to maximize its outputs in order to achieve its performance in the best possible light. As these outputs from stage 1 become inputs to the second stage, high output from stage 1 may lead to poor assessment of performance in the second stage if the optimization criterion at stage 2 is maximization type where more output with less input is preferred. Kao and Hwang (2008) and Liang et al. (2008) overcome this problem under the CRS assumption by assessing that the overall efficiency of the two-stage process as the product of the efficiencies of the two stages. Chen et al. (2009) extend Kao and Hwang (2008) approach by using additive efficiency decomposition under both the CRS and VRS. In the proposed model, we use the VRS assumption, as one of the output variables (average return) used in our empirical application can be negative. The standard VRS DEA model has the translation invariance property so that a constant may be added to all values of the negative valued output variable to make them positive without altering the efficient frontier and the position of the funds relative to the efficient frontier (see Ali and Seiford 1990).

The two stage process proposed in Fig. 7.2 is different from the two-stage process considered in Kao and Hwang (2008), Liang et al. (2008), and Chen et al. (2010) in the sense that we allow new inputs to the second stage in addition to the intermediate measures. The network DEA approach of Färe and Whittaker (1995) and Färe and Grosskopf (1996), the slack-based network DEA approach of Tone and Tsutsui (2009) and the dynamic effects in production networks of Chen (2009) are more general versions of the two-stage process described in Fig. 7.2. However, they do not yield efficiencies at individual stages. We have overcome this problem in the network DEA model used in this chapter. For a review of the relevant recent literature on modeling of network processes, see Cook et al. (2010) and Cook and Zhu (2014). An application of the network DEA approach is available in Lewis and Sexton (2004). The existing approaches cannot be readily adopted to model the situation depicted in Fig. 7.2 and therefore in this study we present a new network DEA approach.

In order to understand the basic concepts behind the proposed two-stage DEA model, consider the simplified version presented in Fig. 7.3. Suppose we have one input ($x1$) to stage 1, one intermediate measure (z), one additional input ($x2$) to stage 2 and one output (y) from stage 2. To measure the overall efficiency of the two-stage process, we first calculate the expected (efficient) output y from stage 2 using input $x1$ indirectly and input $x2$ directly with an intermediate measure z. Assume that the DMU should have produced an output z^* with input $x1$ had it operated efficiently in stage 1 and should have produced an output y^* with inputs z^* and $x2$ in stage 2. Then a measure of overall efficiency is y/y^*, a measure of stage 1 efficiency is z/z^* and a measure of stage 2 efficiency is $(z^* + x2)/(z + x2)$.

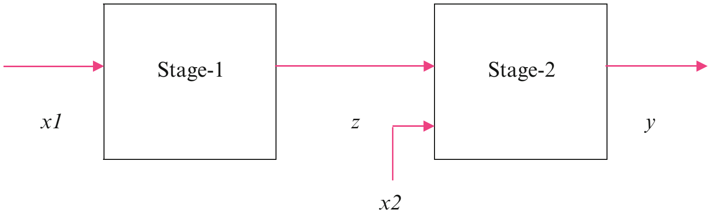


Fig. 7.3 A simplified two-stage framework of mutual fund family performance. This is a simplified version of the complete two-stage DEA model illustrated in Fig. 7.1. $x1$ and $x2$ are the input variables for stage 1 and 2, respectively, and z is the intermediate variable that links the two stages. y is the output variable in stage 2

When calculating the expected (efficient) output of stage 2, we require the intermediate measure to be the expected (efficient) output of stage 1. When this concept is generalized to the case with multiple intermediate measures, the "aggregate" value of intermediate measures must remain the same. According to Liang et al. (2008), such a modeling process treats the two stages as players in a cooperative game where both players "negotiate" on the expected value of intermediate measures. Such a modeling process does not fit into a standard DEA approach. Rather, it optimizes a joint efficiency of the two stages subject to the condition that the intermediate input to stage 2 is the expected output from stage 1. In that regard, the approach used in the two-stage DEA model proposed in this chapter is different from the iterative process used by Holod and Lewis (2011). Their two-stage process is based upon a non-oriented standard DEA model and does not provide separate efficiency estimates for each stage.

Next, we describe the DEA-based procedure used in this chapter to model the relationship between the overall efficiency and the efficiencies at stage 1 and stage 2 in a single mathematical model under the VRS assumption.

Consider a general two-stage DEA network structure for DMU-j with i_1 inputs to stage 1 denoted by $X_j^1 = \left\{x_{1j}^1, x_{2j}^1, \ldots, x_{i_1j}^1\right\}$, i_2 inputs to stage 2 denoted by $X_j^2 = \left\{x_{1j}^2, x_{2j}^2, \ldots, x_{i_2j}^2\right\}$, D intermediate measures denoted by z_{dj} $(d = 1, \ldots, D)$, and s outputs from stage 2 denoted by y_{rj} $(r = 1, \ldots, s)$. With respect to our mutual fund family example in Fig. 7.1, X^1 has two input variables, X^2 has four input variables, z has one variable, and y has one variable. Following Banker et al. (1984), the VRS efficiency score of DMU_o at the first and second stages can be calculated using models (7.1) and (7.2), respectively.

$$
\begin{aligned}
& Max \quad \frac{\sum_d \eta_d^1 z_{do} + u^1}{\sum_{i_1} v_{i_1}^1 x_{i_1 o}^1} \\
& s.t. \quad \frac{\sum_d \eta_d^1 z_{dj} + u^1}{\sum_{i_1} v_{i_1}^1 x_{i_1 j}^1} \leq 1, \quad j = 1, 2, \ldots, n \\
& \qquad v_{i_1}^1, \eta_d^1 \geq \varepsilon; \quad u^1 \text{ free}
\end{aligned} \tag{7.1}
$$

$\left(v_{i_1}^1, \eta_d^1\right)$ are decision variables (weights) associated with the inputs to the first stage and the intermediate measures (outputs from the first stage). u^1 is a free variable associated with returns to scale (RTS) in DEA for stage 1.

$$
\begin{aligned}
& Max \quad \frac{\sum_r u_r y_{ro} + u^2}{\sum_d \eta_d^2 z_{do} + \sum_{i_2} v_{i_2}^2 x_{i_2 o}^2} \\
& s.t. \quad \frac{\sum_r u_r y_{rj} + u^2}{\sum_d \eta_d^2 z_{dj} + \sum_{i_2} v_{i_2}^2 x_{i_2 j}^2} \leq 1, \quad j = 1, 2, \ldots, n \\
& \qquad v_{i_2}^2, u_r, \eta_d^2 \geq \varepsilon; \quad u^2 \text{ free}
\end{aligned} \tag{7.2}
$$

$(v_{i_2}^2, u_r, \eta_d^2)$ are decision variables (weights) associated with the inputs to the second stage, the intermediate measures and outputs from the second stage. u^2 is a free variable associated with RTS in DEA for stage 2.

Note that if we assume $u^1 = u^2 = 0$, then the above models become the CRS models of Charnes et al. (1978) and therefore the following discussion is applicable to the CRS case as well. Similar to Kao and Hwang's (2008) assumption and the centralized model in Liang et al. (2008), we assume that $\eta_d^1 = \eta_d^2 = \eta_d \ (d = 1, \ldots, D)$ in models (7.1) and (7.2). This assumption ensures that in both stages the same multipliers (weights) are applied to the intermediate measures. Then, as far as the intermediate variables are concerned, the expected outputs from stage 1 will be equal to the expected inputs to the second stage.

As in Chen et al. (2009), we compute the overall efficiency as a weighted average of the efficiency scores from stages 1 and 2 as

$$
w_1 \cdot \frac{\sum_d \eta_d z_{do} + u^1}{\sum_{i_1} v_{i_1}^1 x_{i_1 o}^1} + w_2 \cdot \frac{\sum_r u_r y_{ro} + u^2}{\sum_d \eta_d z_{do} + \sum_{i_2} v_{i_2}^2 x_{i_2 o}^2} \tag{7.3}
$$

where w_1 and w_2 are user-specified weights such that $w_1 + w_2 = 1$. If the geometric average as in Kao and Hwang (2008) is used, the product of $\dfrac{\sum_d \eta_d z_{do} + u^1}{\sum_{i_1} v^1_{i_1} x^1_{i_1 o}}$ and $\dfrac{\sum_r u_r y_{ro} + u^2}{\sum_d \eta_d z_{do} + \sum_{i_2} v^2_{i_2} x^2_{i_2 o}}$ will not yield a linear objective function due to the fact that $\sum_d \eta_d z_{do} + \sum_{i_2} v^2_{i_2} x^2_{i_2 o}$ cannot be cancelled. If we assume that $X^2_j = \{\}$ and $u^1 = 0$, the model would reduce to the CRS version and then the approach of Kao and Hwang (2008) can be applied.

In Sect. 7.4.1 we present further details on how the 2-stage model can be generalised by converting (7.3) along with models (7.1) and (7.2) when $\eta^1_d = \eta^2_d = \eta_d$ $(d = 1, \ldots, D)$. We also show how to decompose the overall efficiency and develop a procedure to determine whether the decomposed efficiency scores are unique.

7.4.1 *DEA Model for Two-Stage Network and Efficiency Decomposition*

Since w_1 and w_2 in (7.3) are intended to reflect the relative importance or the contribution of the performance in the first and the second stage to the overall performance, a reasonable choice of weights is the proportion of total resources devoted to each stage. To be more specific, we define

$$\begin{aligned} w_1 &= \frac{\sum_{i_1} v^1_{i_1} x^1_{i_1 o}}{\sum_{i_1} v^1_{i_1} x^1_{i_1 o} + \sum_d \eta_d z_{do} + \sum_{i_2} v^2_{i_2} x^2_{i_2 o}} \quad \text{and} \\ w_2 &= \frac{\sum_d \eta_d z_{do} + \sum_{i_2} v^2_{i_2} x^2_{i_2 o}}{\sum_{i_1} v^1_{i_1} x^1_{i_1 o} + \sum_d \eta_d z_{do} + \sum_{i_2} v^2_{i_2} x^2_{i_2 o}} \end{aligned} \tag{7.4}$$

where $\sum_{i_1} v^1_{i_1} x^1_{i_1 o} + \sum_d \eta_d z_{do} + \sum_{i_2} v^2_{i_2} x^2_{i_2 o}$ represents the total amount of resources (inputs) consumed by the entire two-stage process and $\sum_{i_1} v^1_{i_1} x^1_{i_1 o}$ and $\sum_d \eta_d z_{do} + \sum_{i_2} v^2_{i_2} x^2_{i_2 o}$ represents the amount of resources consumed in the first and the second stage, respectively. These weights are functions of the decision variables of models (7.1) and (7.2).

Hence, under VRS, the overall efficiency score of DMU_o in the two-stage process can be evaluated by solving the following fractional program (7.5). The constraints in (7.5) ensure that the efficiency scores of a DMU in both stages are non-negative and no greater than unity.

$$
\begin{aligned}
\theta_o^* = Max \quad & \frac{\sum_d \eta_d z_{do} + u^1 + \sum_r u_r y_{ro} + u^2}{\sum_{i_1} v_{i_1}^1 x_{i_1 o}^1 + \sum_d \eta_d z_{do} + \sum_{i_2} v_{i_2}^2 x_{i_2 o}^2} \\
s.t. \quad & \frac{\sum_d \eta_d z_{do} + u^1}{\sum_{i_1} v_{i_1}^1 x_{i_1 j}^1} \leq 1, \quad j = 1, 2, \ldots, n \\
& \frac{\sum_r u_r y_{ro} + u^2}{\sum_d \eta_d z_{do} + \sum_{i_2} v_{i_2}^2 x_{i_2 j}^2} \leq 1, \quad j = 1, 2, \ldots, n \\
& 1 \geq w_1 = \frac{\sum_{i_1} v_{i_1}^1 x_{i_1 o}^1}{\sum_{i_1} v_{i_1}^1 x_{i_1 o}^1 + \sum_d \eta_d z_{do} + \sum_{i_2} v_{i_2}^2 x_{i_2 o}^2} \geq w_1^o \\
& 1 \geq w_2 = \frac{\sum_d \eta_d z_{do} + \sum_{i_2} v_{i_2}^2 x_{i_2 o}^2}{\sum_{i_1} v_{i_1}^1 x_{i_1 o}^1 + \sum_d \eta_d z_{do} + \sum_{i_2} v_{i_2}^2 x_{i_2 o}^2} \geq w_2^o \\
& v_{i_1}^1, v_{i_2}^2, u_r, \eta_d \geq \varepsilon, \quad u^1, u^2 \text{ free}
\end{aligned}
\tag{7.5}
$$

Sensitivity analysis of the weights w_1 and w_2 can be performed by adding lower bounds w_1^o and w_2^o on w_1 and w_2. In this study, we substitute 50 % for both w_1^o and w_2^o assuming that operational management and portfolio management are equally important functions.

By applying the Charnes–Cooper transformation, the above fractional programming model (7.5) can be transformed into the following linear programming model (7.6).

$$
\begin{aligned}
\theta_o^* = Max \ & \sum_d \pi_d z_{do} + \sum_r \mu_r y_{ro} + u^A + u^B \\
s.t. \quad & \sum_d \pi_d z_{dj} + u^A \le \sum_{i_1} \omega_{i_1}^1 x_{i_1 j}^1, \quad j = 1, 2, \ldots, n \\
& \sum_r \mu_r y_{rj} + u^B \le \sum_d \pi_d z_{dj} + \sum_{i_2} \omega_{i_2}^2 x_{i_2 j}^2, \quad j = 1, 2, \ldots, n \\
& \sum_{i_1} \omega_{i_1}^1 x_{i_1 o}^1 + \sum_d \pi_d z_{do} + \sum_{i_2} \omega_{i_2}^2 x_{i_2 o}^2 = 1 \\
& 1 \ge \sum_{i_1} \omega_{i_1}^1 x_{i_1 o}^1 \ge w_1^o \\
& 1 \ge \sum_d \pi_d z_{do} + \sum_{i_2} \omega_{i_2}^2 x_{i_2 o}^2 \ge w_2^o \\
& \omega_{i_1}^1, \omega_{i_2}^2, \mu_r, \pi_d \ge \varepsilon, \quad u^A, u^B \text{ free}
\end{aligned}
\tag{7.6}
$$

7.4.1.1 Efficiency Decomposition

Once we obtain an optimal solution to (7.6), the efficiency scores for the two individual stages can be calculated as $\theta_o^{1*} = \dfrac{\sum_d \pi_d^* z_{do} + u^{A*}}{\sum_{i_1} \omega_{i_1}^{1*} x_{i_1 o}^1}$ and $\theta_o^{2*} = \dfrac{\sum_r \mu_r^* y_{ro} + u^{B*}}{\sum_d \pi_d^* z_{do} + \sum_{i_2} \omega_{i_2}^{2*} x_{i_2 o}^2}$. We can also obtain a set of weights as $w_1^* = \sum_{i_1} \omega_{i_1}^{1*} x_{i_1 o}^1$, $w_2^* = 1 - w_1^*$. However, since model (7.6) can have multiple optimal solutions, the θ_o^{1*} and θ_o^{2*} components of overall efficiency may not be unique. Therefore, we follow the procedure adopted by Kao and Hwang (2008) and Chen (2009) to obtain a set of multipliers that would produce the highest first- or second-stage efficiency score while maintaining the overall efficiency score of the entire process fixed. Denote the overall efficiency score of DMU_o obtained by model (7.6) as θ_o^*. We maximize the first-stage efficiency score first while maintaining the overall efficiency score at θ_o^* and the weighted first- and second-stage efficiency scores at no greater than unity as

$$
\begin{aligned}
&\theta_o^{1\,*} = Max \quad \frac{\sum_d \eta_d z_{do} + u^1}{\sum_{i_1} v_{i_1}^1 x_{i_1 o}^1} \\
&\text{s.t. } \frac{\sum_d \eta_d z_{dj} + u^1}{\sum_{i_1} v_{i_1}^1 x_{i_1 j}^1} \le 1, \quad j = 1, 2, \ldots, n \quad \text{(a)} \\
&\frac{\sum_r u_r y_{rj} + u^2}{\sum_d \eta_d z_{dj} + \sum_{i_2} v_{i_2}^2 x_{i_2 j}^2} \le 1, \quad j = 1, 2, \ldots, n \quad \text{(b)} \\
&\frac{\sum_d \eta_d z_{dj} + u^1 + \sum_r u_r y_{ro} + u^2}{\sum_{i_1} v_{i_1}^1 x_{i_1 o}^1 + \sum_d \eta_d z_{do} + \sum_{i_2} v_{i_2}^2 x_{i_2 o}^2} = \theta_o^* \quad \text{(c)} \\
&1 \ge w_1 = \frac{\sum_{i_1} v_{i_1}^1 x_{i_1 o}^1}{\sum_{i_1} v_{i_1}^1 x_{i_1 o}^1 + \sum_d \eta_d z_{do} + \sum_{i_2} v_{i_2}^2 x_{i_2 o}^2} \ge w_1^o \\
&1 \ge w_2 = \frac{\sum_d \eta_d z_{do} + \sum_{i_2} v_{i_2}^2 x_{i_2 o}^2}{\sum_{i_1} v_{i_1}^1 x_{i_1 o}^1 + \sum_d \eta_d z_{do} + \sum_{i_2} v_{i_2}^2 x_{i_2 o}^2} \ge w_2^o \\
&v_{i_1}^1,\, v_{i_2}^2,\, u_r,\, \eta_d \ge \varepsilon, \quad u^1, u^2 \text{ free}
\end{aligned}
\tag{7.7}
$$

In model (7.7), the constraints (a) and (b) ensure that the efficiency scores of all DMUs at both stages are no greater than unity and the constraint (c) maintains the overall efficiency score at θ_o^*. Model (7.7) can be converted into the following equivalent linear program (7.8).

$$
\begin{aligned}
\theta_o^{1\,*} = Max\ & \sum_d \pi_d z_{do} + u^A \\
s.t.\quad & \sum_d \pi_d z_{dj} + u^A \leq \sum_{i_1} \omega_{i_1}^1 x_{i_1 j}^1, \quad j = 1, 2, \ldots, n \\
& \sum_r \mu_r y_{rj} + u^B \leq \sum_d \pi_d z_{dj} + \sum_{i_2} \omega_{i_2}^2 x_{i_2 j}^2, j = 1, 2, \ldots, n \\
& \sum_d \pi_d z_{do} + \sum_r \mu_r y_{rj} + u^A + u^B - \theta_o^* \left(1 + \sum_d \pi_d z_{do} + \sum_{i_2} \omega_{i_2}^2 x_{i_2 o}^2 \right) = 0 \\
& w_1^o \left(1 + \sum_d \pi_d z_{do} + \sum_{i_2} \omega_{i_2}^2 x_{i_2 o}^2 \right) \leq 1 \\
& \left(1 - w_2^o\right) \left(\sum_d \pi_d z_{do} + \sum_{i_2} \omega_{i_2}^2 x_{i_2 o}^2 \right) \geq w_2^o \\
& \sum_{i_1} \omega_{i_1}^1 x_{i_1 o}^1 = 1 \\
& \omega_{i_1}^1, \omega_{i_2}^2, \mu_r, \pi_d \geq \varepsilon, \quad u^A, u^B \ \text{free}
\end{aligned}
\tag{7.8}
$$

Let $\omega_{i_1}^{1\,*}$, $\omega_{i_2}^{2\,*}$, μ_r^*, π_d^*, $u^{A\,*}$, $u^{B\,*}$ represent the optimal values of $\omega_{i_1}^1$, $\omega_{i_2}^2$, μ_r, π_d ,$$ u^A, u^B in model (7.8). Then the first-stage efficiency score is $\theta_o^{1\,*} = \sum_d \pi_d^* z_{do} + u^{A\,*}$ and the optimal weights for the two stages are $w_1^* = \frac{1}{1 + \sum_d \pi_d^* z_{do} + \sum_{i_2} \omega_{i_2}^{2\,*} x_{i_2 o}^2}$ and $w_2^* = 1 - w_1^*$, respectively. The second-stage efficiency score for DMU_o is calculated as $\theta_o^2 = \frac{\theta_o^* - w_1^* \theta_o^{1\,*}}{w_2^*}$. Note that (*) is used in $\theta_o^{1\,*}$ to indicate that the first-stage efficiency score is optimized first. In this case, the resulting efficiency score for the second stage is denoted by θ_o^2 (without *).

Similarly, the following linear program can be formulated to maximize the second-stage efficiency score while maintaining the overall efficiency score at θ_o^* and the weighted first- and second-stage efficiency score at no greater than unity as

$$
\begin{aligned}
\theta_o^{2\,*} = Max\ & \sum_r \mu_r y_{ro} + u^B \\
s.t.\ \ & \sum_d \pi_d z_{dj} + u^A \le \sum_{i_1} \omega_{i_1}^1 x_{i_1 j}^1, \quad j = 1, 2, \ldots, n \\
& \sum_r \mu_r y_{rj} + u^B \le \sum_d \pi_d z_{dj} + \sum_{i_2} \omega_{i_2}^2 x_{i_2 j}^2, \quad j = 1, 2, \ldots, n \\
& \sum_d \pi_d z_{do} + \sum_r \mu_r y_{ro} + u^A + u^B - \theta_o^* \left(\sum_{i_1} \omega_{i_1}^1 x_{i_1 o}^1 + 1 \right) = 0 \\
& \left(1 - w_1^o\right) \sum_{i_1} \omega_{i_1}^1 x_{i_1 o}^1 \ge w_1^o \\
& w_2^o \left(\sum_{i_1} \omega_{i_1}^1 x_{i_1 o}^1 + 1 \right) \le 1 \\
& \sum_d \pi_d z_{do} + \sum_{i_2} \omega_{i_2}^2 x_{i_2 o}^2 = 1 \\
& \omega_{i_1}^1, \omega_{i_2}^2, \ \mu_r, \ \pi_d \ge \varepsilon, \quad u^A, u^B \ \text{free}
\end{aligned}
\tag{7.9}
$$

Let $\omega_{i_1}^{1\,*}, \omega_{i_2}^{2\,*}, \ \mu_r^*, \ \pi_d^*, \ u^{A\,*}, \ u^{B\,*}$ represent the optimal values of $\omega_{i_1}^1, \omega_{i_2}^2, \ \mu_r, \ \pi_d,$ u^A, u^B in model (7.9). Then the second-stage efficiency score is $\theta_o^{2\,*} = \sum_r \mu_r^* y_{r0}$ $+ u^{B\,*}$ and the optimal weights for the two stages are $w_2^* = \dfrac{1}{\sum_{i_1} \omega_{i_1}^{1\,*} x_{i_1 o}^1 + 1}$ and $w_1^* = 1 - w_2^*$, respectively. The first-stage efficiency score is calculated as $\theta_o^1 = \frac{\theta_o^* - w_2^* \theta_o^{2\,*}}{w_1^*}$. If the results satisfy $\theta_o^1 = \theta_o^{1*}$ and $\theta_o^2 = \theta_o^{2*}$, then we may conclude that the decomposed efficiency scores are unique.

As in the conventional DEA models, the efficiency scores obtained for stages 1 and 2 provide information on how an inefficient unit can improve its performance. However, because the optimal (frontier projection) intermediate measures need to be determined, as noted in Chen et al. (2010), one needs to rely on the envelopment form of the DEA model to derive the DEA frontier for the two-stage process. Note that our two-stage network structure is different from the one discussed in Chen et al. (2010) with added additional multiple inputs to the second stage. Therefore, in Sect. 7.4.1.2 we develop a new model for providing information on how to improve the DMUs' performance under our newly developed two-stage DEA network model.

7.4.1.2 Frontier Projection

Model (7.6) does not yield information on optimal intermediate measures. Therefore, following Chen et al. (2010), we develop a model for frontier projection of the DMUs as follows:

$$
\begin{aligned}
&\min\ \left(w_1^*\alpha + w_2^*\beta\right) - \varepsilon\left(\sum_{i_1} s_{i_1}^{1-} + \sum_{i_2} s_{i_2}^{2-} + \sum_{r} s_r^{+}\right)\\
&s.t.\quad \sum_{j=1}^{n}\lambda_j x_{i_1 j}^1 + s_{i_1}^{1-} = \alpha x_{i_1 o}^1, \quad i_1 = 1, 2, \ldots,\\
&\qquad \sum_{j=1}^{n}\lambda_j z_{dj} = \sum_{j=1}^{n}\mu_j z_{hj}, \quad d = 1, 2, \ldots, D\\
&\qquad \sum_{j=1}^{n}\lambda_j = 1\\
&\qquad \sum_{j=1}^{n}\mu_j x_{i_2 j}^2 + s_{i_2}^{2-} = \beta x_{i_2 o}^2, \quad i_2 = 1, 2, \ldots,\\
&\qquad \sum_{j=1}^{n}\mu_j y_{rj} - s_r^{+} = y_{ro}, \quad r = 1, 2, \ldots, s\\
&\qquad \sum_{j=1}^{n}\mu_j = 1\\
&\qquad \lambda_j, \mu_j, s_{i_1}^{1-}, s_{i_2}^{2-}, s_r^{+} \geq 0
\end{aligned}
\tag{7.10}
$$

where w_1^*, w_2^* are obtained from the two-stage network DEA model developed in Sect. 7.4.1.

The above model is based on the production possibility set with $\sum_{j=1}^{n}\lambda_j = 1$ and $\sum_{j=1}^{n}\mu_j = 1$ indicating that both stages exhibit VRS, as in the standard DEA model. $\sum_{j=1}^{n}\lambda_j z_{dj} = \sum_{j=1}^{n}\mu_j z_{hj}$, $d = 1, 2, \ldots, D$ ensures that both stages determine the optimal (frontier projection) intermediate measures.

If we fix α and β in the above model as $\theta_o^{1\,*}$ and $\theta_o^{2\,*}$ obtained from our two-stage model, model (7.10) adopts the principle of the "second-stage" model for calculating DEA slacks (Cooper et al. 2004). In that case, the model becomes

$$
\begin{aligned}
\max\ & \sum_{i_1} s_{i_1}^{1-} + \sum_{i_2} s_{i_2}^{2-} + \sum_{r} s_r^{+} \\
s.t.\ & \sum_{j=1}^{n} \lambda_j x_{i_1 j}^{1} + s_{i_1}^{1-} = \theta_o^{1*} x_{i_1 o}^{1}, \quad i_1 = 1, 2, \ldots, \\
& \sum_{j=1}^{n} \lambda_j z_{dj} = \sum_{j=1}^{n} \mu_j z_{dj}, \quad d = 1, 2, \ldots, D \\
& \sum_{j=1}^{n} \lambda_j = 1 \\
& \sum_{j=1}^{n} \mu_j x_{i_2 j}^{2} + s_{i_2}^{2-} = \theta_o^{2*} x_{i_2 o}^{2}, \quad i_2 = 1, 2, \ldots, \\
& \sum_{j=1}^{n} \mu_j y_{rj} - s_r^{+} = y_{ro}, \quad r = 1, 2, \ldots, s \\
& \sum_{j=1}^{n} \mu_j = 1 \\
& \lambda_j, \mu_j, s_{i_1}^{1-}, s_{i_2}^{2-}, s_r^{+} \geq 0
\end{aligned}
\tag{7.11}
$$

Both stages determine the best projection levels for the intermediate measures as $\sum_{j=1}^{n} \lambda_j^* z_{hj} = \sum_{j=1}^{n} \mu_j^* z_{hj}$. The frontier projection point is given by ($\theta_o^{1*} x_{i_1 o}^{1} - s_{i_1}^{1-*}$, $\sum_{j=1}^{n} \lambda_j^* z_{hj}$, $\theta_o^{2*} x_{i_2 o}^{2} - s_{i_2}^{2-*}$, $y_{ro} + s_r^{+*}$).

7.5 Data and Sampling

The data on US mutual funds are obtained from the Morningstar Direct database. The sample consists of 66 large mutual fund families with total funds under management in each family exceeding \$1 billion USD. The sample period is January 1993 to December 2008 (a total of 1056 family years). The 66 families comprise 1269 individual mutual funds, adding up to 20,304 fund years. For each of these individual funds, we compute monthly return and monthly standard deviation over the 16-year sample period.

Some funds have multiple share classes depending on the fee structure and we consider them as separate mutual funds. Furthermore, we found that some families may offer the same fund to different investors under different names. We treated them as separate funds as well. We included all the funds in the family irrespective of their investment policy or classification, such as money market funds, bond funds, equity funds, and index funds.

During our survey period, some funds may have ceased operations and some funds mostly small funds, do not report all the data that we require. Therefore, we consider only large mutual fund families with total funds under management in each family of at least $1 billion USD. Out of a total of 198 fund families reported in 2008, 101 families (51 %) have a total fund size of at least $1 billion USD. Out of these 101 families, 35 families (34.7 %) are dropped from the study due to non-availability of data on all the input and output variables given in Fig. 7.2.

Our final sample contains 66 mutual fund families. Most of the families that we dropped from the study are small; that is, the fund size of 19 out of the 35 families dropped (54.3 %) is less than $4 billion USD. The two largest families dropped from the study are PIMCO Funds (fund size of $217 billion USD with three mutual funds in it) and Dodge and Cox (fund size of $71 billion USD with three mutual funds in it). Total funds under management in each of the other 14 families dropped from the analysis are between $4 billion USD and $40 billion USD. In DEA, the efficiencies of mutual fund families are assessed relative to the other families in the sample and therefore dropping large families from the sample may affect efficiency scores. However, as only a very small percentage of the dropped funds are large, their impact on the overall assessment is minimal.

Even though the primary focus of this paper is to introduce a novel two-stage DEA model for efficiency decomposition, we make a significant effort to minimize the survivorship bias in the numerical example that we use here to demonstrate the applicability of the proposed model. In mutual fund research, survivorship bias is an important issue. According to Carhart (1997), data used in mutual fund research may often be incomplete due to the following reasons. During the sample period, some funds may have ceased operations or some funds may not report data in poorly performing years. The availability of all the individual fund-level data for the 66 families in our sample throughout the entire survey period implies that all the funds in those selected families are healthy funds and none of them have ceased operations during the survey period.

Summary statistics for the 66 mutual fund families selected in our sample and sorted by total funds under management as of 2008 are presented in Table 7.2. American Funds is by far the largest in terms of funds under management ($1490 billion USD). Vanguard is the next largest with $579 billion USD worth of funds under its control. In our sample, the fund family that offers the greatest number of individual mutual funds is Fidelity Investments, with 94 mutual funds worth $418 billion USD under its management. We consider each mutual fund family in the sample as a separate DMU.

The list of input and output variables used in the DEA model is given in Table 7.3. As illustrated in Fig. 7.2, stage 1 has two inputs and one output and stage 2 has five inputs and one output. These variables are selected following previous studies of mutual fund performance such as Malhotra et al. (2007), Choi and Murthi (2001), Murthi et al. (1997), Nguyen-Thi-Thanh (2006) and Wilkens and Zhu (2005). For each family, the values of the input and output variables are calculated for each year from 1993 to 2008 using the data collected on the individual mutual funds in the family.

Table 7.2 Summary statistics of mutual fund families

Mutual fund family	Number of funds	Total funds (US$)	Average return	Average risk
American funds	42	1,490,594,275,158.00	9.77	2.25
Vanguard	37	579,750,294,615.00	9.16	2.07
Fidelity Investments	94	418,187,641,631.00	10.11	1.56
Franklin Templeton Investments	89	394,375,602,920.00	7.62	0.98
Oppenheimer Funds	48	130,904,879,326.00	7.67	1.88
T. Rowe Price	27	110,222,489,930.00	9.00	3.02
Black Rock	41	108,241,956,534.00	8.02	1.92
Van Kampen	34	82,211,315,521.00	6.81	1.89
Davis Funds	4	63,251,839,275.00	11.09	10.03
Putnam	50	62,171,214,774.00	6.79	1.81
Legg Mason/Western	36	51,461,797,140.00	8.16	2.33
Eaton Vance	22	50,071,998,544.00	7.09	2.32
MFS	41	47,430,409,145.00	7.78	2.22
Lord Abbett	20	47,356,265,200.00	7.91	2.83
Columbia	37	43,501,260,026.00	8.10	1.74
First Eagle	3	38,981,626,920.00	14.04	6.95
Invesco Aim	31	35,566,463,709.00	9.39	3.54
DWS Investments	31	33,964,651,524.00	7.86	2.15
River Source	31	30,546,682,728.00	6.23	1.51
Waddell and Reed	26	27,410,184,204.00	9.79	3.13
Hartford Mutual Funds	12	27,011,372,486.00	10.85	7.13
AllianceBernstein	22	26,226,317,926.00	6.55	14.77
American Century Investments	18	25,473,621,032.00	9.80	5.36
Federated	32	23,816,215,202.00	7.22	2.40
Dreyfus	46	19,419,526,269.00	6.37	1.29
Pioneer Investments	10	18,121,381,148.00	7.30	5.09
Jennison Dryden	13	17,023,214,021.00	8.42	3.66
Nuveen	38	16,070,435,315.00	5.32	0.96
Morgan Stanley	21	16,040,451,102.00	7.57	2.57
Neuberger Berman	8	13,412,797,477.00	10.16	6.09
Calvert	10	12,796,101,648.00	7.12	2.88
Natixis Funds	12	12,672,030,850.00	9.21	4.24
Seligman	18	11,981,528,042.00	11.07	6.03
Principal Funds	12	11,744,149,708.00	7.37	2.26
Main Stay	9	10,363,339,170.00	6.97	3.72
Evergreen	19	10,316,579,030.00	7.77	3.92
Delaware Investments	26	9,854,852,484.00	7.24	2.67
Thrivent	10	9,509,202,204.00	6.94	3.14
Wells Fargo Advantage	16	9,303,483,468.00	8.78	3.72
Victory	7	8,838,386,982.00	9.62	4.95

(continued)

Table 7.2 (continued)

Mutual fund family	Number of funds	Total funds (US$)	Average return	Average risk
Security Funds	6	8,518,518,926.00	9.45	4.13
Selected Funds	2	8,518,518,926.00	9.45	7.48
First American	14	8,032,131,082.00	8.12	2.68
Thornburg	6	5,279,984,262.00	4.77	1.41
First Investors	24	4,859,922,276.00	5.77	1.67
Sentinel	8	4,158,134,728.00	7.73	5.32
Aquila	10	4,081,036,350.00	5.25	1.68
Gabelli	8	3,783,733,120.00	8.55	6.14
JPMorgan	6	3,593,451,011.00	9.51	7.12
Virtus	16	3,110,691,704.00	6.96	3.28
Ariel	2	2,830,081,390.00	8.97	14.88
Baron Capital Group	1	2,622,842,777.00	10.17	21.74
ING Funds	5	2,172,202,625.00	8.49	5.43
Alger	8	2,169,817,261.00	10.61	9.51
RS Funds	4	2,144,384,324.00	12.76	14.88
Merger	1	1,905,360,481.00	7.72	8.47
Pax World	1	1,865,442,450.00	8.06	12.73
Van Eck	2	1,605,535,830.00	9.75	33.45
Transamerica	4	1,581,097,914.00	5.90	4.26
Allianz Funds	4	1,576,090,818.00	11.61	13.61
U.S. Global Investors	5	1,509,204,804.00	9.07	19.79
Allegiant	5	1,228,769,092.00	6.84	3.01
Value Line	9	1,191,193,270.00	8.43	4.79
Eagle Funds	4	1,169,869,484.00	8.91	10.18
Heartland	1	1,131,448,125.00	14.05	30.53
Sun America	10	1,051,166,792.00	7.22	4.42

This table illustrates the summary statistics of the 66 mutual fund families considered in the sample. The sample period is from January 1993 to December 2008. Return on individual mutual funds is obtained from the Morningstar Direct database. Average return is the average monthly return of all individual mutual funds that belong to the family. Average risk is the average of the standard deviations of monthly returns of individual mutual funds that belong to the family. The funds are sorted by total funds under its management at 2008

Summary statistics of the input and output variables are given in Panel A of Table 7.4 and the maximum correlation (Pearson correlation coefficient) between each pair of the variables over the sample period 1993–2008 are given in Panel B. The minimum Pearson correlation coefficient is given in Panel C.

Table 7.3 Input–output variables used in DEA models

Notations	
h_{ij} is the weight defined as investments in fund i as a proportion of total investments in the family j	
N_j is the total number of funds in family j	
Stage 1	
Input variables	
Management fees (I_1):	Is computed as $\sum_{i=1}^{N_j} X_{ij}h_{ij}$, where X_{ij} is the management fee of fund i of family j. This fee includes the fees that are paid out of fund assets to the investment advisors, any other fees payable to the advisors or its affiliates and administrative fees payable to the advisors that are not included in the "other expenses" category
Marketing and distribution fees (I_2) ("12b-1" fees):	Is computed as $\sum_{i=1}^{N_j} Y_{ij}h_{ij}$, where Y_{ij} is the marketing and distribution fees of fund i of family j. This covers the costs of marketing and selling fund shares and sometimes it covers the cost of providing shareholder services
Output variable	
Net asset value (O_1):	Is computed as $\sum_{i=1}^{N_j} P_{ij}h_{ij}$, where P_{ij} is the net asset value of fund i of family j
Stage 2	
Input variables	
Fund size (I_3):	Is computed as $\sum_{i=1}^{N_j} F_{ij}$, where F_{ij} is the total funds in fund i of family j.
Net expense ratio (I_4):	Is computed as $\sum_{i=1}^{N_j} \Psi_{ij}h_{ij}$, where, Ψ_{ij} is the net expense ratio of fund i of family j
Turnover (I_5):	Is computed as $\sum_{i=1}^{N_j} \delta_{ij}h_{ij}$, where δ_{ij} is the turnover ratio of fund i of family j
Standard deviation (I_6):	Is computed as $(A^T A)/N$ where A^T is the transpose of matrix A of excess return over the previous three years and N is the number of observations in the three-year period. For more on this see, Benninga (2008)
Adjusted net asset value (I_7):	Is estimated in the stage 1 DEA model. See Sect. 7.3 for details
Output variable	
Total return (O_2):	Is computed as $\sum_{i=1}^{N_j} h_{ij}r_{ij}$, where, r_{ij} is the annual return of fund i of family j

7.6 Analysis of the Results

In this section we demonstrate application of the two stage DEA model proposed in Sect. 7.4 by examining the relative performance of the US fund families listed in Table 7.2. In Sect. 7.6.1 we analyze the overall performance of the mutual fund

Table 7.4 Summary measures of variables used in the two-stage DEA

Panel A: Summary statistics of DEA input–output variables

	Total funds in fund family (in US billion)		Marketing and distribution fee (%)	Management fees (%)	NAV (in US billion)	Net expense ratio (%)	Turnover ratio	Standard deviation (%)	Return (%)
Maximum	$	1490.59	1.99	1.00	42.29	2.75	1526.94	75.49	113.41
Minimum	$	1.05	0.00	0.16	0.01	0.12	3.00	0.09	−51.21
Mean	$	64.21	0.67	0.51	7.65	0.91	236.66	14.66	12.39
Stddev	$	203.62	0.70	0.33	10.77	0.76	336.01	20.35	29.24

Panel B: Maximum Pearson correlation coefficient between input and output variables (1993–2008)

	Total funds in fund family	Marketing and distribution fee	Management fees	NAV	Net expense ratio	Turnover ratio	Standard deviation	Return
Total funds in family	1.00							
Marketing and distribution	−0.13	1.00						
Management fees	−0.36	0.18	1.00					
NAV	0.83	−0.11	−0.17	1.00				
Net expense ratio	−0.21	0.74	0.63	−0.02	1.00			
Turnover ratio	−0.08	0.22	0.35	−0.04	0.32	1.00		
Standard deviation	−0.11	0.15	0.49	−0.07	0.48	0.15	1.00	
Return	0.11	0.12	0.47	0.22	0.58	0.40	0.68	1.00

(continued)

Table 7.4 (continued)

Panel C: Minimum Pearson correlation coefficient between input and output variables (1993–2008)

	Total funds in fund family	Marketing and distribution fee	Management fees	NAV	Net expense ratio	Turnover ratio	Standard deviation	Return
Total funds in family	1.00							
Marketing and distribution	−0.14	1.00						
Management fees	−0.39	0.04	1.00					
NAV	0.54	−0.29	−0.34	1.00				
Net expense ratio	−0.41	0.53	0.52	−0.55	1.00			
Turnover ratio	−0.16	−0.05	0.00	−0.21	−0.01	1.00		
Standard deviation	−0.22	−0.02	0.31	−0.25	0.19	−0.15	1.00	
Return	−0.06	−0.12	−0.38	−0.18	−0.17	−0.29	−0.57	1.00

Panel A presents the descriptive statistics of the input and output variables used at stage 1 and stage 2 of the two stage DEA model. The maximum and the minimum Pearson correlation coefficient between input and output variables are presented in Panel B and C respectively families using the overall efficiency scores estimated in model (7.5). Thereafter, to gain insights on the source of efficiency/inefficiency of the fund families, we analyze the operational efficiency scores in Sect. 7.6.2 and the portfolio efficiency scores in Sect. 7.6.3.

7.6.1 Overall Efficiency Estimated in the Two-Stage DEA Model

Table 7.5 lists the 16 families that have performed consistently well overall over the most recent 3-year period from 2006 to 2008 based on the overall efficiency estimated in the two-stage DEA model. We judge the consistency of performance of a mutual fund family by the number of times a family has been ranked in the top 2, top 3, and so on up to top 10 during the 3-year period. Since the investigation period is 3 years, the maximum frequency possible under each category is 3.

Vanguard is clearly the best performing fund family over the investigation period (ranked top 2 in all 3 years), followed by Fidelity Investments (ranked top 3 twice), Hartford Mutual Funds (ranked top 4 twice and top 5 three times), Allegiant (ranked top 4 twice and top 6 three times), and American Funds (ranked top 6 twice). It is not surprising that the Vanguard family of funds is the top performer over the most recent 3-year sample period, given its dominance with respect to the market share in terms of funds under passive management (Smith 2010) and adherence to the fund family gospel that low-cost investments deliver the best returns (Dunstan 2012). The Vanguard Group provides the necessary services to run the funds on an at-cost basis (Bogle 2004). As a result, Vanguard has the reputation within the fund management industry as having the lowest operating expenses. In 2008, the Vanguard funds cost, on average, 0.27 % of assets or about 25 % of the industry average (Morningstar 2012). Vanguard is well known among investors for offering mutual funds with the lowest or close to the lowest annual operating expenses and hence the high overall efficiency is not surprising. All the five fund families identified above (Vanguard, Fidelity, Hartford, Allegiant, and American) have substantial market share and a long history averaging over 80 years. Further, they received rankings in the top quartile in the 2007 fund family rankings released by Barron's based on the performance in 2006.

On the other hand, American Century Investments and Neuberger Berman are ranked in the top 2 in one of the 3 years and in the other 2 years both are ranked below 10 showing inconsistency in their performance from 2006 to 2008. The poor performance of Neuberger Berman after 2006 can be linked to the fallout of the global financial crisis.

Now we discuss consistency in the performance over a longer period- the 5 years from 2004 to 2008. Table 7.6 shows the 17 best performing mutual fund families based on overall performance over the 5-year period. As seen in Table 7.6, during this period Vanguard is always ranked in the top 2 and is clearly the best performer.

Table 7.5 Top-performing mutual fund families in the 3-year period from 2006 to 2008 based on their overall efficiency estimated in the two-stage DEA procedure

Fund family	Top 2	Top 3	Top 4	Top 5	Top 6	Top 7	Top 8	Top 9	Top 10
Vanguard	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)
Fidelity Investments	1 (33.3 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	3 (100 %)	3 (100 %)
American Funds	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)
American Century Inv.	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)
Neuberger Berman	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)
Allegiant		1 (33.3 %)	2 (66.7 %)	2 (66.7 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)
Hartford Mutual Funds		1 (33.3 %)	2 (66.7 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)
Dreyfus			1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)
Nuveen				1 (33.3 %)	1 (33.3 %)	2 (66.7 %)	3 (100 %)	3 (100 %)	3 (100 %)
T. Rowe Price				1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	2 (66.7 %)	2 (66.7 %)
Aquila					1 (33.3 %)	2 (66.7 %)	3 (100 %)	3 (100 %)	3 (100 %)
Davis Funds						1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	2 (66.7 %)
First American							1 (33.3 %)	1 (33.3 %)	1 (33.3 %)
Thrivent								1 (33.3 %)	1 (33.3 %)
Franklin Templeton Inv.									1 (33.3 %)
Thornburg									1 (33.3 %)

The table gives the number of times the family has been ranked in the top 2, top 3, etc., over the most recent 3 years from 2006 to 2008. The entry in parentheses gives the percentage of times the family has been ranked under the corresponding category of rankings. For example, Hartford Mutual Funds has been ranked 1, 2, or 3 in only one of the 3 years (33.3 %) from 2006 to 2008

Table 7.6 Top-performing mutual fund families in the 5-year period from 2004 to 2008 based on their overall efficiency estimated in the two-stage DEA procedure

Fund family	Top 2	Top 3	Top 4	Top 5	Top 6	Top 7	Top 8	Top 9	Top 10
Vanguard	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)
Neuberger Berman	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)
Fidelity Investments	1 (20 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	4 (80 %)	4 (80 %)	5 (100 %)	5 (100 %)
American Funds	1 (20 %)	1 (20 %)	1 (20 %)	1 (20 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	4 (80 %)
T. Rowe Price	1 (20 %)	1 (20 %)	1 (20 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	4 (80 %)	4 (80 %)
American Century Inv.	1 (20 %)	1 (20 %)	1 (20 %)	1 (20 %)	1 (20 %)	1 (20 %)	1 (20 %)	1 (20 %)	1 (20 %)
Allegiant		1 (20 %)	2 (40 %)	2 (40 %)	3 (60 %)	3 (60 %)	4 (80 %)	5 (100 %)	5 (100 %)
Hartford Mutual Funds		2 (40 %)	4 (80 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)
Nuveen			1 (20 %)	2 (40 %)	2 (40 %)	4 (80 %)	5 (100 %)	5 (100 %)	5 (100 %)
Dreyfus			1 (20 %)	1 (20 %)	1 (20 %)	1 (20 %)	1 (20 %)	1 (20 %)	1 (20 %)
Columbia				1 (20 %)	1 (20 %)	1 (20 %)	1 (20 %)	1 (20 %)	1 (20 %)
Aquila					2 (40 %)	3 (60 %)	4 (80 %)	5 (100 %)	5 (100 %)
Davis Funds						1 (20 %)	1 (20 %)	1 (20 %)	2 (40 %)
Franklin Templeton Inv.							1 (20 %)	1 (20 %)	3 (60 %)
First American							1 (20 %)	1 (20 %)	1 (20 %)
Thrivent								1 (20 %)	1 (20 %)
Thornburg									1 (20 %)

The table gives the number of times the family has been ranked top 2, top 3, etc., over the most recent five years from 2004 to 2008. The entry in parentheses gives the percentage of times the family has been ranked under the corresponding category of rankings. For example, Allegiant has been ranked 1, 2, or 3 in only one of five years (20 %) from 2004 to 2008

Neuberger Berman is the next best, followed by Fidelity investments, Hartford Mutual Funds and T. Rowe Price. Two out of these five families, Neuberger Berman and T. Rowe Price, do not feature in the list of the five best performers over the most recent 3-year period. The same 16 families reported in Table 7.5 also performed better than the other sampled families over the 5-year period from 2004 to 2008.

Similarly, we investigated the overall performance of the mutual fund families over the 10-year period from 1999 to 2008. The results obtained for the 35 best performing fund families over the ten-year period are presented in Table 7.7. When the window is extended to a longer time horizon, no fund family ranks consistently in the top-10 100 % of the time. The results reveal that Vanguard continues its dominance over the other fund families listed in Tables 7.5 and 7.6 with its performance ranked consistently in the top ten 80 % of the time. The most consistent fund family over the longer term horizon is TransAmerica with 1.5 billion funds under management. TransAmerica which is ranked among the top ten 90 % of the time is considerably smaller in size than Vanguard and as a result is not able to offer a low fee structure in terms of Marketing and Management fees as Vanguard does. However, through effective asset allocation and close attention to its investment mandate, TransAmerica consistently performs well relative to the other fund families in the sample over the 10-year period. Other fund families that demonstrate persistence in overall relative performance in the long term are; Aquila, Sun America, and Barron Capital Group.

One of the main contributions of the proposed two-stage DEA model compared with the conventional DEA models is the decomposition of overall efficiency into two components, namely, operational efficiency and portfolio efficiency. In the next section, we discuss how the fund families have performed over the sample period with respect to operational and portfolio efficiency.

7.6.2 *Operational Management Efficiency*

Table 7.8 lists the 13 fund families that perform relatively better from 2006 to 2008 based on the operational efficiency scores estimated in the proposed two-stage DEA model. The operational efficiency score reflects how well a fund family has managed its resources in securing or generating funds for that family. Here, we observe that three families have been ranked top 2 in all 3 years of assessment; Vanguard, T. Rowe Price, and American Century Investments. According to the overall efficiency score rankings reported in Table 7.5, only Vanguard performs at this level. The next-best performer under operational efficiency is Neuberger Berman, with rankings of 3 or better in all 3 years, followed by American Funds and Fidelity Investments.

The top-performing families in terms of operational efficiency over the 5-year period 2004–2008 reported in Table 7.9 reveal that the same 13 families reported in Table 7.8 also performed better than the other sampled families over this 5-year

Table 7.7 Overall consistent performance of mutual fund families in the 10-year period from 1993 to 2008

Fund family	Top-1	Top-2	Top-3	Top-4	Top-5	Top-6	Top-7	Top-8	Top-9	Top-10
Van Eck	4 (40 %)	4 (40 %)	5 (50 %)	5 (50 %)	5 (50 %)	5 (50 %)	5 (50 %)	5 (50 %)	5 (50 %)	5 (50 %)
U.S. Global Investors	4 (40 %)	4 (40 %)	4 (40 %)	4 (40 %)	4 (40 %)	5 (50 %)	5 (50 %)	5 (50 %)	5 (50 %)	5 (50 %)
First Eagle	2 (20 %)	2 (20 %)	2 (20 %)	2 (20 %)	3 (30 %)	3 (30 %)	4 (40 %)	4 (40 %)	4 (40 %)	4 (40 %)
Eaton Vance	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)
Invesco Aim	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)
Waddell and Reed	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)
Nuveen	1 (10 %)	2 (20 %)	2 (20 %)	2 (20 %)	3 (30 %)	4 (40 %)	4 (40 %)	4 (40 %)	5 (50 %)	5 (50 %)
Thornburg	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	2 (20 %)	2 (20 %)
Ariel	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)
Baron Capital Group	1 (10 %)	2 (20 %)	3 (30 %)	3 (30 %)	3 (30 %)	3 (30 %)	4 (40 %)	5 (50 %)	6 (60 %)	6 (60 %)
RS Funds	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)
Heartland	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	2 (20 %)	2 (20 %)	3 (30 %)	3 (30 %)	3 (30 %)
Sun America	1 (10 %)	2 (20 %)	2 (20 %)	2 (20 %)	3 (30 %)	3 (30 %)	4 (40 %)	5 (50 %)	6 (60 %)	6 (60 %)
Transamerica		2 (20 %)	5 (50 %)	6 (60 %)	9 (90 %)	9 (90 %)	9 (90 %)	9 (90 %)	9 (90 %)	9 (90 %)
Vanguard		1 (10 %)	2 (20 %)	6 (60 %)	6 (60 %)	7 (70 %)	7 (70 %)	7 (70 %)	7 (70 %)	8 (80 %)
Davis Funds		1 (10 %)	2 (20 %)	2 (20 %)	2 (20 %)	2 (20 %)	2 (20 %)	2 (20 %)	2 (20 %)	3 (30 %)
JennisonDryden		1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)
Principal Funds		1 (10 %)	2 (20 %)	2 (20 %)	2 (20 %)	3 (30 %)	3 (30 %)	3 (30 %)	3 (30 %)	5 (50 %)
Alger		1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)
Thrivent			1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)
Aquila			1 (10 %)	1 (10 %)	2 (20 %)	4 (40 %)	5 (50 %)	7 (70 %)	8 (80 %)	8 (80 %)
Eagle Funds				3 (30 %)	4 (40 %)	4 (40 %)	5 (50 %)	5 (50 %)	5 (50 %)	5 (50 %)
Fidelity Invest				1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)	1 (10 %)
Security Funds				1 (10 %)	1 (10 %)	2 (20 %)	3 (30 %)	3 (30 %)	4 (40 %)	4 (40 %)
ING Funds					2 (20 %)	2 (20 %)	3 (30 %)	3 (30 %)	4 (40 %)	5 (50 %)

(continued)

Table 7.7 (continued)

Fund family	Top-1	Top-2	Top-3	Top-4	Top-5	Top-6	Top-7	Top-8	Top-9	Top-10
Franklin Templeton Inv						1 (10 %)	2 (20 %)	3 (30 %)	3 (30 %)	3 (30 %)
Allegiant						1 (10 %)	1 (10 %)	1 (10 %)	2 (20 %)	3 (30 %)
Merger							1 (10 %)	1 (10 %)	2 (20 %)	2 (20 %)
Value Line							1 (10 %)	3 (30 %)	3 (30 %)	3 (30 %)
American Century Invst								1 (10 %)	1 (10 %)	1 (10 %)
MainStay								1 (10 %)	1 (10 %)	1 (10 %)
Columbia									1 (10 %)	1 (10 %)
Hartford Mutual Funds										1 (10 %)
AllianceBernstein										1 (10 %)
Dreyfus and Virtus										1 (10 %)

This table illustrates the number of times a particular family has been ranked as top-1, top-2, top -3, . . ., top-10, over the most recent 10 years (1999–2008). The percentage of times over 10 years is given in the parenthesis. These figures illustrate whether a fund family has been performing well consistently over the most recent 10 years

Table 7.8 Top-performing mutual fund families in the 3-year period from 2006 to 2008 based on their operational efficiency estimated in the proposed two-stage DEA procedure

Fund family	Top 2	Top 3	Top 4	Top 5	Top 6	Top 7	Top 8	Top 9	Top 10
Vanguard	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)
T. Rowe Price	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)
American Century Inv.	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)
Neuberger Berman	2 (66.7 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)
American Funds	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	3 (100 %)	3 (100 %)
Fidelity Investments		2 (66.7 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)
Wells Fargo Advantage			2 (66.7 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)
Allegiant				2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	3 (100 %)	3 (100 %)
Dreyfus					2 (66.7 %)	2 (66.7 %)	3 (100 %)	3 (100 %)	3 (100 %)
Hartford Mutual Funds						3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)
Victory							2 (66.7 %)	2 (66.7 %)	3 (100 %)
Columbia								1 (33.3 %)	2 (66.7 %)
Nuveen									1 (33.3 %)

The table gives the number of times the family has been ranked in the top 2, top 3, etc., over the 3 years from 2006 to 2008. The entry in parentheses gives the percentage of times the family has been ranked under the corresponding category of rankings. For example, Fidelity Investments has been ranked 1, 2, or 3 in only two of the three years (66.7 %) from 2006 to 2008

Table 7.9 Top-performing mutual fund families in the 5-year period from 2004 to 2008 based on their operational efficiency estimated in the proposed two-stage DEA procedure

Fund family	Top 2	Top 3	Top 4	Top 5	Top 6	Top 7	Top 8	Top 9	Top 10
American Century Inv.	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)
T. Rowe Price	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)
Vanguard	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)
Neuberger Berman	3 (60 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)
American Funds	1 (20 %)	1 (20 %)	1 (20 %)	1 (20 %)	2 (40 %)	2 (40 %)	3 (60 %)	4 (80 %)	5 (100 %)
Wells Fargo Advantage	1 (20 %)	1 (20 %)	4 (80 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)
Fidelity Investments		3	4 (80 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)
Hartford Mutual Funds			1 (20 %)	1 (20 %)	2 (40 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)
Allegiant				3 (60 %)	3 (60 %)	4 (80 %)	4 (80 %)	5 (100 %)	5 (100 %)
Dreyfus					3 (60 %)	3 (60 %)	5 (100 %)	5 (100 %)	5 (100 %)
Victory						1 (20 %)	3 (60 %)	4 (80 %)	5 (100 %)
Columbia								1 (20 %)	3 (60 %)
Nuveen								1 (20 %)	2 (40 %)

The table gives the number of times the family has been ranked in the top 2, top 3, etc., over the 5 years from 2004 to 2008. The entry in parentheses gives the percentage of times the family has been ranked under the corresponding category of rankings. For example, Hartford Mutual Funds has been ranked 1, 2, 3 or 4 in only one of 5 years (20 %) from 2004 to 2008 period. The top 5 performers from 2006 to 2008 are also the top 5 performers over the 5-year period. Once again, when the window is extended to reflect the long term nature of investing (1999–2008) as shown in Table 7.10, the same mutual fund families continue to demonstrate their comparative advantage with low fee structures. Not surprisingly, the two fund families that are ranked 1 or 2 in any calendar year on the basis of operational efficiency throughout the 10-year window; Vanguard (100 %) and T. Rowe Price (100 %) are both among the top 6 in terms of funds under management having portfolios exceeding 100 billion USD. The fee structure of these two fund families reveal that they are able to keep the costs significantly below industry average. This is evident especially in the case of marketing fees where Vanguard and T. Rowe Price both have fees less than 2.3 % compared to the industry average of 44.4 %.

Seven additional fund families join Vanguard and T. Rowe Price by consistently outperforming the other families in terms of operational efficiency. The continually dominating top 10 families are; American Century Investments, Wells Fargo Advantage, American funds, Neuberger Berman, Fidelity investments, Allegiant and Dreyfus. All these fund families have realized high levels of net asset values given their levels of management and marketing fees. We were not able to obtain data on variables such as salaries and rent that may be relevant for operational performance assessment. If it were possible, one could easily include them in the model to further improve the discriminatory power of mutual fund families based on their operational performance.

7.6.3 Portfolio Management Efficiency

Portfolio management efficiency measures how well a mutual fund family manages its investment portfolio to realize high returns subject to a chosen set of factors that may influence returns. Portfolio efficiency is important information not only for investors in making their investment decisions but also for fund family administrators in assessing the performance of their portfolio managers. The fund family administrators may be able to judge how well their fund managers have performed relative to their competitors using the proposed portfolio management efficiency score (measure). The benefits of the proposed efficiency measure do not stop there. Relative performance at the portfolio management level is vital information for recruiting agencies to identify the best-performing fund managers and those who are underperforming.

As in the previous cases, Tables 7.11 and 7.12 lists the fund families that have been ranked at or above different levels of ranking in the last three- and 5-year periods respectively based on portfolio efficiency. According to Table 7.11, Hartford Mutual Funds, Vanguard, Nuveen, Aquila, Davis Funds and Sun America have managed their portfolios relatively better securing a rank of at least 2 during the 3-year period beginning 2006. High performance in stage 2 implies that the mutual fund family has gained relatively high returns with their existing level of fund size,

Table 7.10 Top-performing mutual fund families in the 10-year period from 1999 to 2008 based on their operational efficiency estimated in the proposed two-stage DEA procedure

Fund family	Top-1	Top-2	Top-3	Top-4	Top-5	Top-6	Top-7	Top-8	Top-9	Top-10
T. Rowe Price	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)
Vanguard	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)
American Century Inv	5 (50 %)	9 (90 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)
Wells Fargo Advantage	5 (50 %)	6 (60 %)	6 (60 %)	9 (90 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)
American Funds	3 (30 %)	3 (30 %)	3 (30 %)	4 (40 %)	4 (40 %)	5 (50 %)	5 (50 %)	7 (70 %)	9 (90 %)	10 (100 %)
Neuberger Berman		5 (50 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)
Fidelity Investments			3 (30 %)	5 (50 %)	9 (90 %)	9 (90 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)
Hartford Mutual Funds			1 (10 %)	4 (40 %)	4 (40 %)	5 (50 %)	8 (80 %)	8 (80 %)	8 (80 %)	8 (80 %)
Allegiant				1 (10 %)	6 (60 %)	8 (80 %)	9 (90 %)	10 (100 %)	10 (100 %)	10 (100 %)
Dreyfus						6 (60 %)	8 (80 %)	10 (100 %)	10 (100 %)	10 (100 %)
Nuveen							2 (20 %)	5 (50 %)	6 (60 %)	7 (70 %)
Victory							1 (10 %)	3 (30 %)	4 (40 %)	5 (50 %)
Columbia									2 (20 %)	8 (80 %)
JP Morgan									1 (10 %)	2 (20 %)

The table gives the number of times the family has been ranked in the top 1, top 2, top 3, etc., over the 10 years from 1999 to 2008. The entry in parentheses gives the percentage of times the family has been ranked under the corresponding category of rankings. For example, Vanguard and T. Rowe Price have been ranked first equal in terms of operational efficiency in all years between 1999 and 2008 (100 %)

Table 7.11 Top-performing mutual fund families in the 3-year period from 2006 to 2008 based on their portfolio efficiency estimated in the proposed two-stage DEA procedure

Fund family	Top 2	Top 3	Top 4	Top 5	Top 6	Top 7	Top 8	Top 9	Top 10
Hartford Mutual Funds	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)
Vanguard	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)
Nuveen	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)
Aquila	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)
Davis Funds	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)
Sun America	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)
Principal Funds	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)	3 (100 %)
Van Eck	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)
Fidelity Investments	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)
American Funds	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)
Thornburg	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)
Baron Capital Group	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)
Evergreen	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)
Jennison Dryden	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)
Security Funds	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)
Selected Funds	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)
Transamerica	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)
U.S. Global Investors	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)
Gabelli	1 (33.3 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)
Allegiant	1 (33.3 %)	1 (33.3 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)
ING Funds	1 (33.3 %)	1 (33.3 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)
Natixis Funds	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)
Franklin Templeton Inv.	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)
First Eagle	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)

(continued)

Table 7.11 (continued)

Fund family	Top 2	Top 3	Top 4	Top 5	Top 6	Top 7	Top 8	Top 9	Top 10
Eagle Funds	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)	2 (66.7 %)
American Century Inv.	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)
Dreyfus	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)
Neuberger Berman	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)
First American	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)
Thrivent	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)
Alger	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)	1 (33.3 %)

The table gives the number of times the family has been ranked in the top 2, top 3, etc., over the 3 years from 2006 to 2008. The entry in parentheses gives the percentage of times the family has been ranked under the corresponding category of rankings

Table 7.12 Top-performing mutual fund families in the 5-year period from 2004 to 2008 based on their portfolio efficiency estimated in the proposed two-stage DEA procedure

Fund family	Top 2	Top 3	Top 4	Top 5	Top 6	Top 7	Top 8	Top 9	Top 10
Hartford Mutual Funds	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)
Vanguard	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)
Nuveen	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)
Aquila	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)
Davis Funds	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)
Principal Funds	4 (80 %)	4 (80 %)	4 (80 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)	5 (100 %)
Sun America	4 (80 %)	4 (80 %)	4 (80 %)	4 (80 %)	4 (80 %)	4 (80 %)	4 (80 %)	5 (100 %)	5 (100 %)
Baron Capital Group	4 (80 %)	4 (80 %)	4 (80 %)	4 (80 %)	4 (80 %)	4 (80 %)	4 (80 %)	4 (80 %)	4 (80 %)
Jennison Dryden	4 (80 %)	4 (80 %)	4 (80 %)	4 (80 %)	4 (80 %)	4 (80 %)	4 (80 %)	4 (80 %)	4 (80 %)
Security Funds	4 (80 %)	4 (80 %)	4 (80 %)	4 (80 %)	4 (80 %)	4 (80 %)	4 (80 %)	4 (80 %)	4 (80 %)
Selected Funds	4 (80 %)	4 (80 %)	4 (80 %)	4 (80 %)	4 (80 %)	4 (80 %)	4 (80 %)	4 (80 %)	4 (80 %)
American Funds	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	4 (80 %)	4 (80 %)
Franklin Templeton Inv.	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	4 (80 %)	4 (80 %)	4 (80 %)	4 (80 %)	4 (80 %)
First Eagle	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	4 (80 %)	4 (80 %)	4 (80 %)	4 (80 %)	4 (80 %)
Transamerica	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	4 (80 %)	4 (80 %)	4 (80 %)	4 (80 %)	4 (80 %)
Van Eck	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	4 (80 %)
Fidelity Investments	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)
Neuberger Berman	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)
U.S. Global Investors	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)
First American	2 (40 %)	2 (40 %)	2 (40 %)	2 (40 %)	2 (40 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)
ING Funds	2 (40 %)	2 (40 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)
Thornburg	2 (40 %)	2 (40 %)	2 (40 %)	2 (40 %)	2 (40 %)	2 (40 %)	2 (40 %)	2 (40 %)	2 (40 %)
Thrivent	2 (40 %)	2 (40 %)	2 (40 %)	2 (40 %)	2 (40 %)	2 (40 %)	2 (40 %)	2 (40 %)	2 (40 %)
Evergreen	2 (40 %)	2 (40 %)	2 (40 %)	2 (40 %)	2 (40 %)	2 (40 %)	2 (40 %)	2 (40 %)	2 (40 %)

(continued)

Table 7.12 (continued)

Fund family	Top 2	Top 3	Top 4	Top 5	Top 6	Top 7	Top 8	Top 9	Top 10
Gabelli	1 (20 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)
Van Kampen	1 (20 %)	2 (40 %)	2 (40 %)	2 (40 %)	2 (40 %)	2 (40 %)	3 (60 %)	3 (60 %)	3 (60 %)
T. Rowe Price	1 (20 %)	1 (20 %)	2 (40 %)	2 (40 %)	2 (40 %)	2 (40 %)	2 (40 %)	4 (80 %)	4 (80 %)
Allegiant	1 (20 %)	1 (20 %)	2 (40 %)	2 (40 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)	3 (60 %)
Columbia	1 (20 %)	1 (20 %)	1 (20 %)	2 (40 %)	2 (40 %)	2 (40 %)	2 (40 %)	2 (40 %)	2 (40 %)
Black Rock	1 (20 %)	1 (20 %)	1 (20 %)	2 (40 %)	2 (40 %)	2 (40 %)	2 (40 %)	2 (40 %)	2 (40 %)
Eagle Funds	1 (20 %)	1 (20 %)	1 (20 %)	1 (20 %)	1 (20 %)	2 (40 %)	2 (40 %)	2 (40 %)	2 (40 %)

The table gives the number of times the family has been ranked in the top 2, top 3, etc., over the 5 years from 2004 to 2008. The entry in parentheses gives the percentage of times the family has been ranked under the corresponding category of rankings

transaction costs (net expense ratio), turnover ratio, risk exposure (standard deviation) and net asset value. The next-best set of mutual fund families in Table 7.11 includes Principal Funds, Van Eck, Fidelity Investments, American Funds, Thornburg, Baron Capital Group, Evergreen, Jennison Dryden, Security Funds, Selected Funds, Transamerica and US Global Investors. Under the portfolio efficiency measure, the top 5 performers from 2006 to 2008 (see Table 7.11) are also the top 5 performers over the 5-year period from 2004 to 2008 (see Table 7.12).

Fiduciary Insight, using Morningstar Direct data, produces quarterly research reports on the major managed fund families in the US. These reports rank fund families by the percentage of individual funds within the family that have either "passed" the fiduciary score or the "appropriate" classification. The ranking of fund families based on our stage 2 portfolio efficiency scores and the ranking by Fiduciary Insight (Fiduciary Insight 360 2009) for the period ending December 2008 are remarkably similar. Fiduciary Insight reports that Aquila, American Funds, American Century Investments, Baron Capital Group, Eagle Funds, Franklin Templeton Investments and Vanguard belong to the top quartile of the funds as of December 31, 2008. In this, we observe that the traditional approaches used by the fund family ranking organizations may rely only on portfolio efficiency rather than on an overall efficiency measure that covers both the operational management and portfolio management aspects of performance. In recognition of performance over the 3-year period from 2008 to 2010, Transamerica received four 2008 Lipper Fund awards. Transamerica also received for the eighth consecutive year dating back to 2001, the DALBAR Mutual Fund Service Award for excellence in customer service. However, according to Table 7.12, Transamerica is not one of the top performers. A plausible reason for the differences in the rankings of some families, such as Transamerica, based on the overall efficiency scores estimated in the two-stage DEA model and those offered by family ranking organizations may be that these organizations consider a small sample of fund families that satisfy specific investment criteria. Their selection criteria may also vary from time to time.

Table 7.13 presents a summary of the results for long term performance over the window 1999–2008. Table 7.13 only presents fund families that rate in the top 10 at least 60 % of the time. Sixty out of the 66 fund families are ranked within the top 10 at some stage of the 10-year period beginning 1999. The families that are not ranked at least once in the top 10 in any given calendar year are; Seligman (top ranking 13th in 1999), Calvert (top ranking 14th in 2002), Victory (top ranking 16th in 2003), Morgan Stanley (top ranking 19th in 2006), Wells Fargo Advantage (top ranking 22nd in 2000) and Allianz Funds (top ranking 35th in 1999). In terms of portfolio performance over the long run, Vanguard, Nuveen and Aquila are the best.

Table 7.14 provides the rankings of individual fund families each year from 1993 to 2008 based on the overall, operational, and portfolio efficiencies estimated in the two-stage DEA model. We report only the top 10 mutual fund families listed in Table 7.14 to conserve space. It is clear in Table 7.14 that the overall efficiency of mutual fund families may be affected by their portfolio and operational efficiencies being at varying degrees. For example, Vanguard is both operationally and portfolio efficient with a rank of 1 and hence is overall efficient throughout the period

Table 7.13 Top-performing mutual fund families in the 10-year period from 1999 to 2008 based on their portfolio efficiency estimated in the proposed two-stage DEA procedure

Mutual fund family	Top-1	Top-2	Top-3	Top-4	Top-5	Top-6	Top-7	Top-8	Top-9	Top-10
Vanguard	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)
Nuveen	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)
Aquila	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)
Principal Funds	7 (70 %)	9 (90 %)	9 (90 %)	9 (90 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)
Davis Funds	6 (60 %)	7 (70 %)	8 (80 %)	8 (80 %)	8 (80 %)	9 (90 %)	10 (100 %)	10 (100 %)	10 (100 %)	10 (100 %)
Franklin Templeton Inv	8 (80 %)	8 (80 %)	8 (80 %)	8 (80 %)	8 (80 %)	9 (90 %)	9 (90 %)	9 (90 %)	9 (90 %)	9 (90 %)
Transamerica	8 (80 %)	8 (80 %)	8 (80 %)	8 (80 %)	8 (80 %)	9 (90 %)	9 (90 %)	9 (90 %)	9 (90 %)	9 (90 %)
Baron Capital Group	7 (70 %)	7 (70 %)	7 (70 %)	7 (70 %)	8 (80 %)	8 (80 %)	8 (80 %)	8 (80 %)	8 (80 %)	8 (80 %)
Selected Funds	7 (70 %)	7 (70 %)	7 (70 %)	7 (70 %)	7 (70 %)	7 (70 %)	7 (70 %)	7 (70 %)	7 (70 %)	7 (70 %)
U.S. Global Investors	7 (70 %)	7 (70 %)	7 (70 %)	7 (70 %)	7 (70 %)	7 (70 %)	7 (70 %)	7 (70 %)	7 (70 %)	7 (70 %)
Security Funds	6 (60 %)	7 (70 %)	7 (70 %)	7 (70 %)	7 (70 %)	7 (70 %)	7 (70 %)	7 (70 %)	7 (70 %)	7 (70 %)
Hartford MFs	6 (60 %)	7 (70 %)	7 (70 %)	7 (70 %)	7 (70 %)	7 (70 %)	7 (70 %)	7 (70 %)	7 (70 %)	7 (70 %)
American Funds	5 (50 %)	5 (50 %)	5 (50 %)	6 (60 %)	6 (60 %)	6 (60 %)	6 (60 %)	6 (60 %)	7 (70 %)	7 (70 %)
Sun America	5 (50 %)	5 (50 %)	5 (50 %)	5 (50 %)	5 (50 %)	5 (50 %)	5 (50 %)	6 (60 %)	7 (70 %)	7 (70 %)
Thornburg	5 (50 %)	6 (60 %)	6 (60 %)	6 (60 %)	6 (60 %)	6 (60 %)	6 (60 %)	6 (60 %)	6 (60 %)	7 (70 %)
First American	3 (30 %)	4 (40 %)	4 (40 %)	4 (40 %)	4 (40 %)	4 (40 %)	6 (60 %)	6 (60 %)	6 (60 %)	7 (70 %)
Van Eck	5 (50 %)	5 (50 %)	5 (50 %)	5 (50 %)	5 (50 %)	5 (50 %)	5 (50 %)	6 (60 %)	6 (60 %)	7 (70 %)
ING Funds	5 (50 %)	5 (50 %)	5 (50 %)	6 (60 %)	6 (60 %)	6 (60 %)	6 (60 %)	6 (60 %)	6 (60 %)	6 (60 %)
First Eagle	5 (50 %)	5 (50 %)	5 (50 %)	5 (50 %)	5 (50 %)	6 (60 %)	6 (60 %)	6 (60 %)	6 (60 %)	6 (60 %)
Eagle funds	3 (30 %)	3 (30 %)	3 (30 %)	3 (30 %)	3 (30 %)	4 (40 %)	5 (50 %)	6 (60 %)	6 (60 %)	6 (60 %)
T. Rowe Price	3 (30 %)	3 (30 %)	3 (30 %)	4 (40 %)	4 (40 %)	4 (40 %)	4 (40 %)	4 (40 %)	6 (60 %)	6 (60 %)

The table gives the number of times the family has been ranked in the top 2, top 3, etc., over the 5 years from 1999 to 2008. The entry in parentheses gives the percentage of times the family has been ranked under the corresponding category of rankings

Table 7.14 Rankings of a sample of mutual fund families in each year from 1993 to 2008 based on the efficiency estimated in the proposed two-stage DEA procedure

Mutual fund family	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Ranking based on overall efficiency																
Allegiant	24	5	33	12	9	21	42	33	32	8	7	9	8	6	4	3
American Century Inv.	52	45	48	52	44	1	8	43	44	27	9	20	24	19	1	32
American Funds	17	3	2	1	1	1	2	1	5	22	12	10	6	2	6	19
Dreyfus	19	32	6	2	3	4	4	3	25	5	5	11	12	4	14	14
Fidelity Investments	48	20	3	4	2	2	23	2	33	24	4	7	3	9	3	2
Hartford Mutual Funds	59	54	38	35	51	32	20	59	29	14	3	3	4	3	5	4
Neuberger Berman	30	27	19	9	38	49	48	41	36	25	2	2	2	12	2	17
Nuveen	1	4	5	6	5	3	3	4	2	2	6	4	7	8	7	5
T. Rowe Price	9	30	30	3	34	10	1	37	39	20	1	1	5	5	9	12
Vanguard	43	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Ranking based on operational efficiency																
Allegiant	1	6	7	5	5	4	5	4	6	6	5	5	7	9	5	5
American Century Inv.	41	2	2	2	4	1	1	3	2	2	1	1	2	2	1	1
American Funds	4	8	5	1	1	1	4	1	9	1	8	8	10	6	9	1
Dreyfus	19	5	6	6	6	5	6	6	7	7	6	6	8	8	6	6
Fidelity Investments	15	7	8	7	7	6	7	5	5	5	4	3	5	4	3	3
Hartford Mutual Funds	56	52	50	48	49	48	39	38	4	4	3	4	6	7	7	7
Neuberger Berman	35	3	4	4	3	3	3	2	3	3	2	2	3	3	2	2
Nuveen	1	10	10	9	9	7	8	7	8	8	7	9	11	13	11	10
T. Rowe Price	22	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Vanguard	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Ranking based on Portfolio efficiency																
Allegiant	41	6	24	20	18	25	38	30	29	9	17	14	6	12	1	4

(continued)

Table 7.14 (continued)

Mutual fund family	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
American Century Inv.	36	45	30	39	40	1	22	40	46	43	28	37	36	41	1	46
American Funds	18	2	1	1	1	1	1	1	4	29	13	9	1	1	1	35
Dreyfus	7	31	12	5	7	11	7	1	21	6	4	15	11	1	20	26
Fidelity Investments	43	14	1	1	1	1	23	1	32	30	1	17	1	27	1	1
Hartford Mutual Funds	45	32	1	1	32	1	1	42	27	15	1	1	1	1	1	2
Neuberger Berman	14	22	23	23	39	36	43	39	38	40	3	1	1	28	1	33
Nuveen	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
T. Rowe Price	1	34	28	9	37	22	1	35	40	28	1	1	4	9	9	25
Vanguard	49	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

The fund families listed here are the top 10 best-performing families under the overall efficiency estimated in the proposed two-stage DEA model. See Table 7.5 for a summary of their performance in different subsample periods. To conserve space, we list only the top 10 families

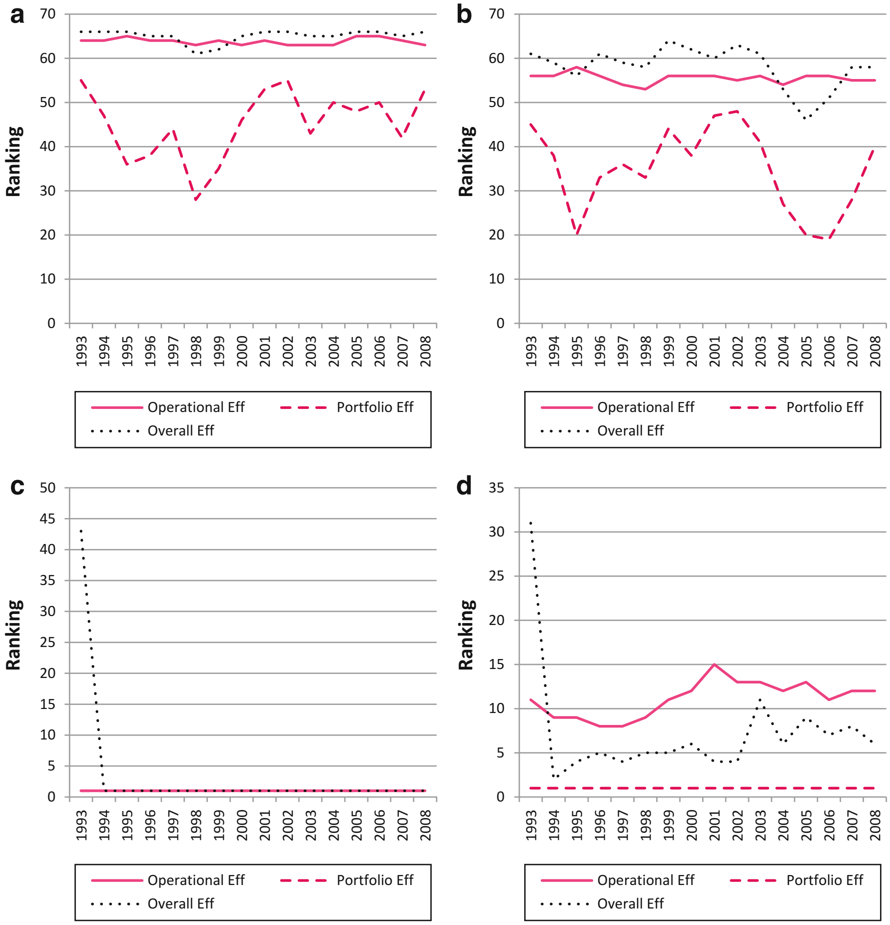


Fig. 7.4 Impact of operational efficiency and portfolio efficiency on the overall efficiency for Allianz, Morgan Stanley, Vanguard and Aquila fund families. This figure illustrates how the rankings of Allianz Funds, Morgan Stanley, Vanguard, and Aquila fund families based on their overall, operational and portfolio efficiency change over the period 1993–2008. (**a**) Allianz Funds. (**b**) Morgan Stanley. (**c**) Vanguard. (**d**) Aquila

1994–2008. T. Rowe Price, on the other hand, is operationally efficient during the period 1994–2008 maintaining a rank of 1. However, T. Rowe Price is not portfolio efficient (except in 2003 and 2004) and therefore is not overall efficient in most of the years. More on the effect of portfolio and operational efficiencies on overall efficiency for a set of fund families is discussed and illustrated graphically in the next section.

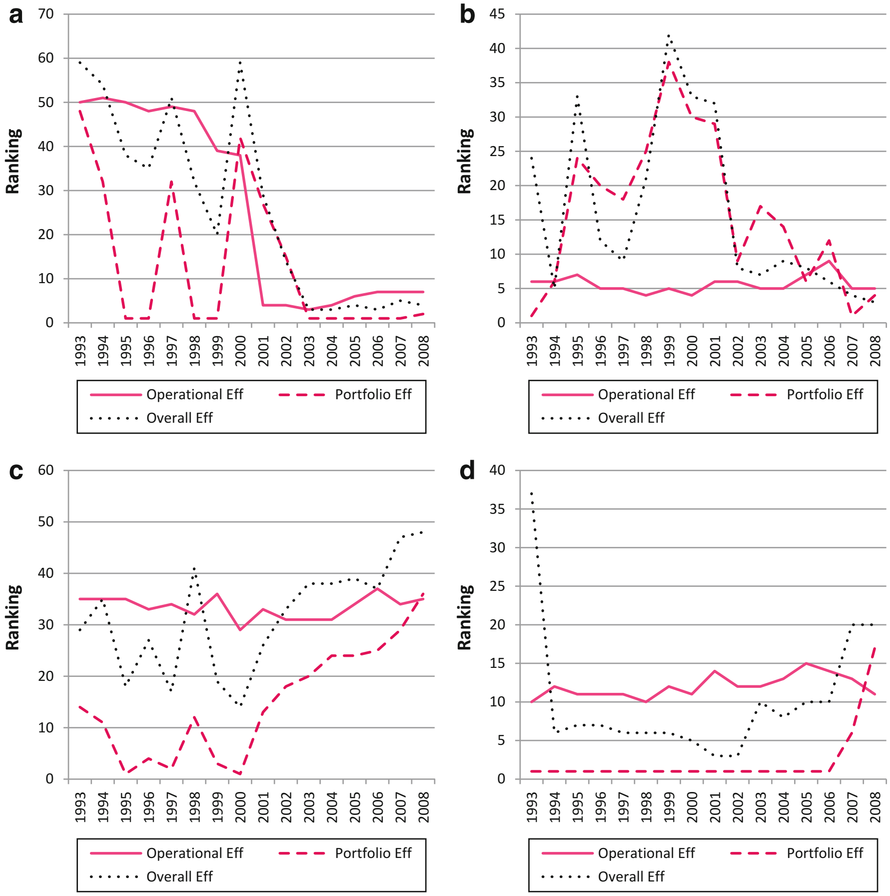


Fig. 7.5 Impact of operational efficiency and portfolio efficiency on the overall efficiency for Hartford, Allegiant, Putnam and Franklin Templeton fund families. This figure illustrates how the rankings of Hartford Mutual Fund, Allegiant, Putnam, and Franklin Templeton Investments on their overall, operational and portfolio efficiency change over the period 1993–2008. (**a**) Hartford Mutual Fund. (**b**) Allegiant. (**c**) Putnam. (**d**) Franklin Templeton Investments

7.6.4 Variation in Efficiency Across Time and Fund Families

Selecting a few families as examples, we now illustrate graphically how operational efficiency and portfolio efficiency may affect the overall efficiency of fund families over time. Panels (a) and (b) of Fig. 7.4 give the graphs for Allianz Funds and Morgan Stanley, respectively. Both these funds perform consistently poorly overall, due to consistent poor operational and portfolio performance. Panel (c) shows that Vanguard's continual overall performance is due to the excellent performance in both the operational and portfolio fronts. The graph in panel (d) of Fig. 7.4 for

Aquila indicates that the reason for its continued good overall performance is mainly due to the consistency in its portfolio management efficiency. Panels (a) and (b) of Fig. 7.5 show the corresponding graphs for Hartford Mutual Funds and Allegiant. These are examples of fund families that have improved their performance after 2003. The improvement of Hartford Mutual Funds family after 2003 is mainly due to the improvement in operational and portfolio efficiencies, and in the case of Allegiant more or less due to the improvement in portfolio efficiency. On the other hand, Putnam and Franklin Templeton Investments, whose graphs are shown in Panels (c) and (d), respectively, in Fig. 7.5, reveal that the poor portfolio efficiency appears to be the main contributor to their declining overall performance towards the end of the sample period.

Four families (Allianz Funds, Morgan Stanley, Hartford Mutual Fund, and Putnam), illustrated in Figs. 7.4 and 7.5, show relatively poor portfolio performance in 1993. We notice similar performance in several other mutual fund families in the sample as well. This is clear evidence of the effect of the 1991 currency crisis on the portfolios managed by some mutual fund families. The improvement shown in the relative rankings after 1994 suggests quick recovery from the crisis in 1991. Vanguard and Aquila have managed their mutual funds relatively efficiently during all financial crisis periods from 1993 to 2008.

During the last two quarters of 1990 and the first quarter of 1991, the US economy experienced a sustained period of negative growth. Other significant shocks to the market during the sample period include the collapse of Long-Term Capital Management in 1998, the dotcom bubble and the subsequent market crash in March 2000, the market meltdown following the September 11 attacks in New York and the Enron debacle, and the recent global financial crisis (GFC) that impacted the markets post July 2007. The effects of the GFC continued well into the years that followed. In Figs. 7.4 and 7.5, we observe that the portfolio efficiency of Allianz Funds, Morgan Stanley, and Putnam families have been seriously affected (low portfolio efficiency ranking) by the recessions of 1990–1991 and 2000–2002 and the fallout from the GFC over the period 2007–2009. These three fund families have high exposure investment across domestic and international equity markets: Allianz Funds (94 %), Morgan Stanley (79 %), and Putnam (68 %). In contrast, even though Hartford Mutual Fund has been affected by the downturn in market activity in 1991 and 2000 to an extent similar to that of the three aforementioned fund families, it has not been affected as much by the problems resulting from the GFC in 2007. The better showing of Hartford Mutual Fund in the later period may be attributed to improved operational and portfolio efficiencies in part driven by an appropriate fee structure. The performance of Allegiant has been affected by the 1998 Long-Term Capital Management collapse and the 2000 recession and has survived the impact of the 2007 crisis. The standout fund family within our sample, Vanguard, as far as operational, portfolio, and overall efficiencies are concerned, has been exceptional throughout the full sample period.

Aquila performs extremely well in terms of portfolio efficiency, but due to its poor operational efficiency its overall efficiency is also low. The Franklin

Templeton Investments family has done extremely well in its portfolio management until 2006. As far as operational efficiency is concerned, it has not done well, with a rank of around 10. The operational and portfolio management performance of the Hartford Mutual Fund family is not relatively satisfactory up to 2003, but has shown tremendous improvement in these areas thereafter. The Allegiant family's operational efficiency is relatively satisfactory over the sample period, but its portfolio efficiency is relatively weak. However, Allegiant's overall performance shows an improvement after 2005. Allianz Funds and Morgan Stanley show inferior overall performance due to their poor performance in both operational and portfolio management areas and show no sign of improvement over the sample period. The Putnam family's operational management performance is relatively poor throughout the sample period. Its portfolio management performance has been relatively satisfactory until 2000 and has deteriorated thereafter. Overall, the above analysis clearly shows that the proposed DEA model is able to capture the dynamics of the operational and portfolio management efficiencies and overall efficiency of mutual fund families.

7.6.5 Frontier Projection of DMUs

Another important feature of DEA is its ability to provide information to make inefficient DMUs efficient. In this subsection, we demonstrate this feature in a selected set of mutual fund families. Such information is very important for a fund family's management decision making.

In Sect. 7.4.1.2, following Chen et al. (2010) we develop a model for frontier projection of mutual fund families deemed inefficient according to the proposed two-stage DEA model. We apply the frontier projection model with the values of the input, output and intermediate variables corresponding to the year 2008. The input, output and intermediate variable changes required for making the inefficient mutual fund families efficient are illustrated in Table 7.15 for a selected set of families. Under the column "NAV" (the intermediate measure), a positive percentage indicates that NAV should be increased, and a negative percentage indicates that NAV should be decreased in order to make the fund family efficient. Positive values with respect to the other input variables in Table 7.15 indicate that they should be decreased by the corresponding percentages.

According to Table 7.15, no changes are required for any of the input (I_1, I_2, I_3, I_4, $I_{5,}$ and I_6), output (O_2) and intermediate (NAV) variables of Vanguard and Fidelity Investments, as they are operational, portfolio and overall efficient in year 2008. This observation tallies with the 2008 ranking of these two families in Table 7.14, where they are ranked within the top three as far as overall, operational, and portfolio efficiencies are concerned. The percentage changes of the variables in the second stage for Davis Funds are all zero, indicating that it is portfolio efficient in 2008. This is evident in Tables 7.11 and 7.12, where this family has been ranked within the top 2 during the period 2004–2008. However, Davis Funds is operationally inefficient and

Table 7.15 Frontier projections of mutual fund families

Mutual fund family	Marketing and distribution fees (I_2)	Management fees (I_1)	NAV	Fund size (I_3)	Standard deviation (I_6)	Net expense ratio (I_4)	Turnover (I_5)	Average return (O_2)
American Funds	72 %	8 %	37 %	74 %	24 %	24 %	24 %	18
Vanguard	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0
Fidelity Investments	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0
Franklin Templeton Investments	29 %	29 %	24 %	72 %	19 %	19 %	19 %	14
Davis Funds	51 %	40 %	0 %	0 %	0 %	0 %	0 %	0
Hartford Mutual Funds	24 %	24 %	0 %	0 %	0 %	0 %	0 %	0
American Century Investments	0 %	0 %	0 %	34 %	34 %	34 %	52 %	4
Oppenheimer Funds	42 %	42 %	−80 %	35 %	35 %	35 %	35 %	23
Morgan Stanley	62 %	43 %	16 %	40 %	40 %	40 %	50 %	19
Neuberger Berman	0 %	0 %	0 %	30 %	40 %	30 %	30 %	30
Principal Funds	35 %	35 %	0 %	0 %	0 %	0 %	0 %	0
Security Funds	53 %	42 %	86 %	19 %	18 %	28 %	18 %	36
Baron Capital Group	66 %	66 %	0 %	0 %	0 %	0 %	0 %	0
ING Funds	31 %	31 %	0 %	0 %	0 %	0 %	0 %	0
RS Funds	46 %	46 %	0 %	0 %	0 %	0 %	0 %	0
Merger	65 %	65 %	0 %	0 %	0 %	0 %	0 %	0
Pax World	53 %	38 %	88 %	15 %	15 %	24 %	15 %	3
Van Eck	51 %	51 %	0 %	0 %	0 %	0 %	0 %	0
U.S. Global Investors	58 %	58 %	50 %	21 %	21 %	34 %	21 %	14
Eagle Funds	36 %	35 %	0 %	0 %	0 %	0 %	0 %	0
Heartland	55 %	55 %	0 %	0 %	0 %	0 %	0 %	0

This table gives the percentage changes required in the inputs, outputs, and intermediate measure (NAV) of the two stages illustrated in Fig. 7.1 in order to make the inefficient mutual fund families efficient. For input variables, the percentages represent a decrease and for output variables they represent an increase. In order to improve the efficiency of stage 1 and stage 2, the input variables (I_1 to I_6) should be decreased by the percentages given, the intermediate measure (NAV) should be changed by the percentage given under the column labeled NAV, and the output should be increased by the percentage given under (O_2)

therefore it has to decrease the marketing and distribution fee and the management fee by 51 and 40 % in the first stage, respectively, to become operationally efficient. These changes will make Davis Funds overall efficient as well. Evidence presented in Tables 7.5 and 7.6 supports this finding, as Davis Funds appears at the bottom of these tables as far as overall efficiency is concerned. On the other hand, American Century Investments is operationally efficient but not efficient in managing the portfolio. This family needs to increase its return by 4 % and decrease its inputs at stage 2—I_3, I_4, I_5, and I_6 by the following percentages: 34 %, 34 %, 52 %, and 34 %, respectively—in order to become portfolio efficient and thereby become overall efficient. According to the entries in Table 7.15, Morgan Stanley is a poor performer in 2008 with inefficient operational and portfolio management. This is evident in Fig. 7.4b with the overall, operational and portfolio rankings of this family lying in the range 20–60. For Morgan Stanley to be overall efficient, it needs to reduce all its inputs at stage 1 and stage 2 by the percentages given in Table 7.15 and increase its stage 1 output or the intermediate measure (NAV) by 16 % and increase the return (output) by 19 %. On the other hand, Oppenheimer Funds may decrease all its inputs at stages 1 and 2 (and NAV) by the percentages given in Table 7.15 and increase its return by 23 % in order to become efficient. These two examples (Morgan Stanley with a positive change in NAV and Oppenheimer Funds with a negative change in NAV) demonstrate an interesting feature of the proposed DEA model; that is, the model treats the intermediate variable, NAV, as both an input as well as an output. In the proposed DEA model, the optimal NAV is determined by both stages through coordination in such a way that the performances of both stages are maximized. Technique such as a stochastic frontier approach cannot treat a variable as an input and as an output within the same model.

7.7 Concluding Remarks

The main objective of this chapter is to present a two-stage network DEA model and demonstrate its application by assessing the relative performance of large mutual fund families in the US. It is well documented that the mutual fund industry in the US is the largest such industry in the world and its well being is important to a strong global economy. Hence a heightened understanding at both the operational level and portfolio performance level of fund families is of importance as we move forward to a time of increased numbers in retirement relying upon their investment income for day-to-day living costs. Unlike traditional performance measures such as the Sharpe, Treynor and Sortino measures, the DEA model proposed in this chapter allows a combination of several factors of performance such as; returns, fees and charges, risk of investment, stock selection style, portfolio management skills and operational management skills into a single measure in evaluating the overall performance of a mutual fund family relative to the other families included in the sample.

The presented two-stage DEA model provides greater insight into the performance of mutual fund families by decomposing the overall efficiency into two components: operational efficiency and portfolio efficiency. In addition to mutual fund families, the proposed DEA model can also be applied to other financial institutions such as banks, insurance companies, credit unions, etc.

The performance of the mutual fund families assessed over the period 1993–2008 using the proposed DEA model reveals that the two-stage model is able to highlight those mutual fund families that may have managed their portfolios well during financial crisis periods as well as which of the two components; operational management and portfolio management may have been the contributory factor for their superior/inferior performance. This is useful information as it can aid individual and institutional investors when making investment decisions and also enables administrators of fund families to judge how well their portfolio managers have performed relative to their competitors.

References

Alexakis P, Tsolas I (2011) Appraisal of mutual equity fund performance using data envelopment analysis. Multinat Finance J 15:273–296

Ali AI, Seiford LM (1990) Translation invariance in data envelopment analysis. Oper Res Lett 9:403–405

Anderson R, Brockman C, Giannikos C, McLeod R (2004) A nonparametric examination of real-estate mutual fund efficiency. Int J Bus Econ 3:225–238

Andersen P, Petersen, N (1993) A procedure for ranking efficient units in data envelopment analysis. Manag Sci 39:1261–1264

Banker RD, Charnes A, Cooper WW (1984) Some models for the estimation of technical and scale inefficiencies in data envelopment analysis. Manag Sci 30:1078–1092

Banker R, Cooper W, Seiford L, Thrall R, Zhu J (2004) Returns to scale in different DEA models. Eur J Oper Res 154:345–362

Basso A, Funari S (2001) A data envelopment analysis approach to measure the mutual fund performance. CEJOR 135(3):477–492

Basso A, Funari S (2003) Measuring the performance of ethical mutual funds: a DEA approach. J Oper Res Soc 54(5):521–531

Basso A, Funari S (2005) A generalized performance attribution technique for mutual funds. CEJOR 13(1):65–84

Benninga S (2008) Financial modeling, 3rd edn. The MIT Press, Cambridge, MA

Bogle J (2004) Re-mutualizing the mutual fund industry—the alpha and the omega. Boston College Law Rev 45:391–422

Brown S, Goetzmann W, Park J (2001) Careers and survival: competition and risk in the hedge fund and CTA industry. J Financ 56:1869–1886

Carhart M (1997) Persistence in mutual fund performance. J Financ 52:57–82

Chang K (2004) Evaluating mutual fund performance: an application of minimum convex input requirements set approach. Comput Oper Res 31:929–940

Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision-making units. Eur J Oper Res 2:429–444

Charnes A, Cooper W, Seiford L, Stutz J (1982) A multiplicative model for efficiency analysis. Socioecon Plann Sci 16:223–224

Chen CM (2009) A network-DEA model with new efficiency measures to incorporate the dynamic effect in production networks. Eur J Oper Res 194:687–699

Chen Z, Lin R (2006) Mutual fund performance evaluation using data envelopment analysis with new risk measures. OR Spectr 28:375–398
Chen Y, Cook WD, Li N, Zhu J (2009) Additive efficiency decomposition in two-stage DEA. Eur J Oper Res 196:1170–1176
Chen Y, Cook WD, Zhu J (2010) Deriving the DEA frontier for two-stage processes. Eur J Oper Res 202:138–142
Chen Y, Chis Y, Li M (2011) Mutual fund performance evaluation—application of system BCC model. S Afr J Econ 79:1–16
Chilingerian J, Sherman HD (2004) Health care applications: from hospitals to physician, from productive efficiency to quality frontiers. In: Cooper WW, Seiford LM, Zhu J (eds) Handbook on data envelopment analysis. Springer, Boston
Choi YK, Murthi BPS (2001) Relative performance evaluation of mutual funds: a non-parametric approach. J Bus Finance Acc 28:853–876
Cook WD, Zhu J (2014) Data envelopment analysis—a handbook of modeling internal structures and networks. Springer, New York
Cook WD, Liang L, Zhu J (2010) Measuring performance of two-stage network structures by DEA: a review and future perspective. Omega 38:423–430
Cook WD, Tone K, Zhu J (2014) Data envelopment analysis: prior to choosing a model. Omega 44:1–4
Cooper WW, Seiford LM, Zhu J (2004) Handbook on data envelopment analysis. Kluwer Academic, Boston
Cooper WW, Seiford LM, Tone K (2006) Introduction to data envelopment analysis and its uses. Springer, New York
Daraio C, Simar L (2006) A robust nonparametric approach to evaluate and explain the performance of mutual funds. Eur J Oper Res 175:516–542
Drake L, Weyman-Jones T (1996) Productive and allocative inefficiencies in UK Building societies: a comparison of non-parametric and stochastic frontier techniques. Manch Sch 64 (1):22–37
Dunstan B (2012, February 15) Price customers into the market. The Australian Financial Review, p 32
Eling M (2006) Performance measurement of hedge funds using data envelopment analysis. Fin Mkts Portfolio Mgmt 20(4):442–471
Elton E, Gruber M, Blake C (2006) The adequacy of investment choices offered by 401 K plans. J Public Econ 90:303–318
Elton E, Gruber MJ, Green TC (2007) The impact of mutual fund family membership on investor risk. J Financ Quant Anal 42:257–278
Färe R, Grosskopf S (1996) Productivity and intermediate products: a frontier approach. Econ Lett 50:65–70
Färe R, Whittaker G (1995) An intermediate input model of dairy production using complex survey data. J Agric Econ 46:201–213
Fare R, Grosskopf S, Grifell-Tatje E, Knox-Lovell C (1997) Biased technical change and the Malmquist productivity index. Scand J Econ 99(1):119–127
Farrell MJ (1957) The management of productive efficiency. J R Stat Soc Ser A 120:253–290
Favero C, Papi L (1995) Technical efficiency and scale efficiency in the Italian banking sector: a non-parametric approach. Appl Econ 27(4):385–395
Ferris S, Chance D (1987) The effects of 12b-1 plans on mutual fund expense ratios: a note. J Financ 42:1077–1082
Fiduciary Insight 360 (2009) Fund family fiduciary rankings. Data as of December 31, 2008. Document retrieved July 12, 2010. http://www.fi360.com/press/pdfs/rankings.pdf
Galagedera D, Silvapulle P (2002) Australian mutual fund performance appraisal using data envelopment analysis. Manag Financ 28(9):60–73
Gregoriou G (2003) Performance appraisal of funds of hedge funds using data envelopment analysis. J Wealth Manag 5:88–95

Gregoriou G, Sedzro K, Zhu J (2005) Hedge fund performance appraisal using data envelopment analysis. Eur J Oper Res 164:555–571
Haslem J, Scheraga C (2003) Data envelopment analysis of Morningstar's large-cap mutual funds. J Invest 12(4):41–48
Haslem J, Scheraga CA (2006) Data envelopment analysis of Morningstar's small-cap mutual funds. J Invest 12:87–92
Holod D, Lewis HF (2011) Resolving the deposit dilemma: a new DEA bank efficiency model. J Bank Financ 35:2801–2810
Hsu C, Lin J (2007) Mutual fund performance and persistence ion Taiwan: a non parametric approach. Serv Ind J 27:509–523
Hu J, Chang T (2008) Decomposition of mutual fund performance. Appl Financ Econ Lett 4:363–367
Investment Company Institute (ICI) (2010) Investment company fact book, 50th edn. Washington, DC
Investment Company Institute (ICI) (2013) Investment company fact book. 53rd edn. Washington, DC
Kao C, Hwang SN (2008) Efficiency decomposition in two-stage data envelopment analysis: an application to non-life insurance companies in Taiwan. Eur J Oper Res 185:418–429
Kempf A, Ruenzi S (2008) Family matters: rankings within fund families and fund inflows. J Bus Finance Acc 35:177–199
Koopmans T (1951) An analysis of production as an efficient combination of activities. In: Koopmans TC (ed) Activity analysis of production and allocation. Wiley, New York, pp 33–97
Latzko D (1999) Economies of scale in mutual fund administration. J Financ Res 22:331–340
Lewis H, Sexton T (2004) Network DEA: efficiency analysis of organizations with complex internal structure. Comput Oper Res 31:1365–1410
Liang L, Cook WD, Zhu J (2008) DEA models for two-stage processes: game approach and efficiency decomposition. Nav Res Logist 55:643–653
Lozano S, Guiterrez E (2008) Data envelopment analysis of mutual funds based on second-order stochastic dominance. Eur J Oper Res 189:230–244
Malhotra D, McLeod R (1997) An empirical analysis of mutual fund expenses. J Financ Res 20:175–190
Malhotra DK, Martin R, Russel P (2007) Determinants of cost efficiencies in the mutual fund industry. Rev Financ Econ 16:323–334
Matallin C, Soler A, Tortosa-Ausina E (2014) On the informativeness of persistence for evaluating mutual fund performance using partial frontiers. Omega 42(1):47–64
McLeod R, Malhotra D (1994) A re-examination of the effects of 12b-1 plans on mutual fund expense ratios. J Financ Res 17:231–240
McMullen P, Strong R (1998) Selection of mutual funds using data envelopment analysis. J Bus Econ Stud 4:1–12
Morningstar Research Pty Ltd (2012) Vanguard data pages. Document retrieved April 3, 2012. http://www.morningstar.com/FundFamily/vanguard.html
Murthi BP, Choi Y, Desai P (1997) Efficiency of mutual funds and portfolio performance measurement: a non-parametric approach. Eur J Oper Res 98:408–418
Nguyen-Thi-Thanh H (2006) On the use of data envelopment analysis in hedge fund selection. Working paper, Université d'Orléans
Powers J, McMullen P (2000) Using data envelopment analysis to select efficient large market cap securities. J Bus Manag 7:31–42
Premachandra I, Powel J, Shi J (1998) Measuring the relative efficiency of fund management strategies in New Zealand using a spreadsheet-based stochastic data envelopment analysis model. Omega 26:319–331
Premachandra I, Bhabra G, Sueyoshi T (2009) DEA as a tool for bankruptcy assessment: a comparative study with logistic regression technique. Eur J Oper Res 193:412–424

Premachandra IM, Zhu J, Watson J, Galagedera DUA (2012) Best-performing US mutual fund families from 1993 to 2008: evidence from a novel two-stage DEA model for efficiency decomposition. J Bank Finance 36:3302–3317

Rubio J, Hassan M, Merdad H (2012) Non-parametric performance measurement of international and Islamic mutual funds. Account Res J 25(3):208–226

Sedzro K, Sardano D (2000) Mutual fund performance evaluation using data envelopment analysis. In: Dahiya SB (ed) The current state of business disciplines. Spellbound, Rohtak, pp 1125–1144

Sengupta J (2003) Efficiency tests for mutual fund portfolios. Appl Financ Econ 13:869–876

Siems TF, Barr RS (1998) Benchmarking the productive efficiency of U.S. banks. Financial industry studies. Federal Reserve Bank of Dallas, pp 11–24

Smith DM (2010) The economics of mutual funds. In: Haslem JA (ed) Mutual funds: portfolio structures, analysis, management, and stewardship. Wiley, Hoboken

Tone L, Tsutsui M (2009) Network DEA: a slacks-based measure approach. Eur J Oper Res 197:243–252

Tower E, Zheng W (2008) Ranking of mutual fund families: minimum expenses and maximum loads as markers for moral turpitude. Int Rev Econ 55:315–350

Wilkens K, Zhu J (2005) Classifying hedge funds using data envelopment analysis. In: Gregoriou GN, Rouah F, Karavas VN (eds) Hedge funds: strategies, risk assessment, and returns. Beard Books, Washington, DC

Zhao X, Yue W (2012) A multi-subsystem fuzzy DEA model with its application in mutual funds management companies' competence evaluation. Procedia Comput Sci 1(1):2469–2478

Zhu J (2002) Quantitative models for performance evaluation and benchmarking: data envelopment analysis with spreadsheets. Kluwer, Boston

Chapter 8
DEA Performance Assessment of Mutual Funds

Antonella Basso and Stefania Funari

Abstract The objectives of this paper are manyfold. First we present a comprehensive review of the literature of DEA models for the performance assessment of mutual funds. Then we discuss the problem of the presence of negative returns in DEA modeling for mutual funds and we identify a DEA model that is financially justified and tackles the issue of negative returns in a natural way. Moreover, we present an empirical application on real market data, considering different risk measures. We consider also different holding periods, which include both a period of financial crisis and one of financial recovery. Moreover, we compare the results of the DEA performance measure with those obtained with traditional financial indicators.

Keywords DEA • Mutual fund performance evaluation • Negative data • Sharpe index • Sortino index

8.1 Introduction

The applications of data envelopment analysis (DEA) to the assessment of the performance of mutual funds have become more and more numerous in the last years. If we consider the applications to conventional mutual funds, socially responsible investment (SRI) mutual funds, Islamic funds, pension funds, exchange-traded funds (ETFs), hedge funds, commodity trading advisors (CTAs) and managed future funds, the number of papers published on international journals and books totals about 100 (data referred to the end of 2014).

However, if we look at the performance indicators used in the financial practice to compare mutual funds on the basis of their historical results, we find that the

A. Basso (✉)
Department of Economics, Ca' Foscari University of Venice,
Cannaregio 873, 30121 Venice, Italy
e-mail: basso@unive.it

S. Funari
Department of Management, Ca' Foscari University of Venice,
Cannaregio 873, 30121 Venice, Italy
e-mail: funari@unive.it

J. Zhu (ed.), *Data Envelopment Analysis*, International Series in Operations
Research & Management Science 238, DOI 10.1007/978-1-4899-7684-0_8

indicators most often used do not include any of the performance scores obtained with a DEA model.

We may wonder what is the reason of this lack of "financial visibility" with both practitioners and big data providers working in the financial market. In our opinion, we may find two main reasons.

The first explanation is connected to the difficulty of providing a clear financial interpretation of the DEA indicators which can be plainly grasped by financial professionals.

A second possible explanation lies in the relative sophistication of the DEA models, compared to some of the traditional financial indicators, such as the Sharpe, Treynor and Sortino ratios. This is especially true for many of the advanced DEA models proposed in the more recent literature.

We have to raise an issue that often affects the financial data on mutual funds and can cause a drawback in DEA modeling: the presence of negative mean returns, which is often observed, especially in periods of financial crisis, and requires the usage of special devices in the formulation of the DEA models.

The main objectives of this paper are manyfold and may be summarized as follows.

First, we present a comprehensive review of the literature of DEA models for the evaluation of the performance of mutual funds (Sect. 8.2).

Secondly, we discuss the problem of the presence of negative returns in the DEA modeling for mutual funds (Sect. 8.3).

Thirdly, we identify a DEA model that is financially well justified and tackles the issue of negative returns in a natural way, being inspired by financial considerations rather than derived from mathematical technicalities. At the same time, it relies on one of the basic DEA models, so that it is relatively simple to implement (Sect. 8.4).

In Sect. 8.5 we outline some of the traditional indicators that are most widely used in finance to evaluate the mutual fund performance; they also suffer from serious drawbacks in presence of negative mean returns.

A fourth objective of this contribution is to present an empirical application of the DEA model chosen on real market data, considering different risk measures. The empirical investigation considers also different holding periods and is carried out both on a period of financial crisis and on a period of financial recovery. Moreover, we compare the results of the DEA performance measure with those obtained with the traditional financial indicators (Sect. 8.6).

8.2 DEA Literature on Mutual Funds

The issue of the assessment of the performance of mutual funds was not among the first applications of the DEA methodology. Indeed, the first contributions which proposed a DEA model to study the performance of mutual funds were published only a little more than 15 years ago. Among the pioneering papers we find Murthi et al. (1997), McMullen and Strong (1998), Basso and Funari (2001), Choi and Murthi (2001), Tarim and Karan (2001) and Galagadera and Silvapulle (2002).

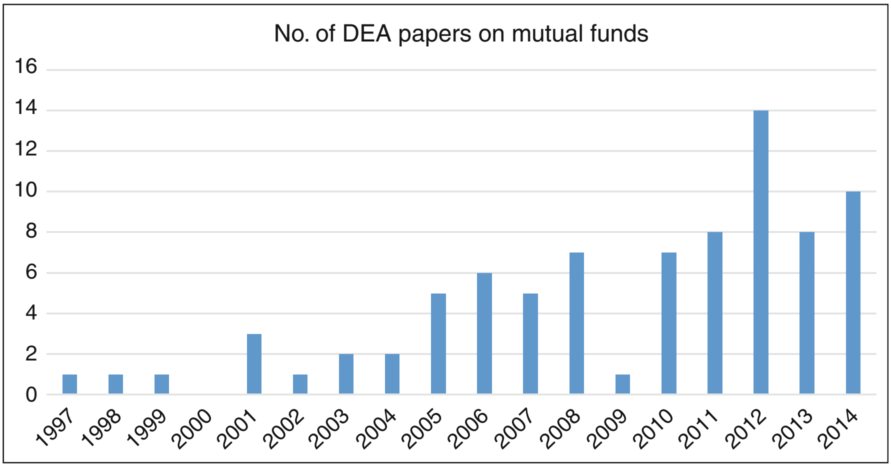


Fig. 8.1 Number of published papers on DEA models for the evaluation of the performance of mutual funds (including SRI funds and Islamic funds), pension funds and ETFs by year of publication

The number of papers published yearly on this subject begins to be considerable starting from 2003, and becomes copious in the more recent years. The situation is summarized in Fig. 8.1, which illustrates the number of papers published in scientific journals or books that adopt a DEA model for the evaluation of the performance of mutual funds by year of publication. The number of publications considered in Fig. 8.1 includes the papers which focus on traditional mutual funds, socially responsible investment (SRI) mutual funds and Islamic funds, but also pension funds and exchange-traded funds (ETFs).

Along the same line, Fig. 8.2 displays the number of published papers that use a DEA approach for the assessment of the performance of hedge funds, commodity trading advisors (CTAs) and managed futures funds.

As can be seen from Figs. 8.1 and 8.2, the overall number of papers on DEA assessment of mutual funds published up to now is quite relevant. Table 8.1 presents the various contributions in detail, reporting for each paper the kind of DEA models proposed/used, a brief summary of the main features of the study and the main characteristics of the empirical analysis carried out (the geographical area and time period of the data and the number of funds considered).

Analogously, Table 8.2 presents the papers on the DEA performance evaluation of socially responsible investment (SRI) mutual funds, which exploits the ability of DEA to take into account not only the financial features but also a measure of the degree of social responsibility of mutual funds.

Moreover, in the more recent years the DEA approach has been applied also to a special kind of SRI funds: the funds that follow the rules of Islamic finance (Shariah compliant); for a review of the contributions on Islamic mutual funds see Table 8.3.

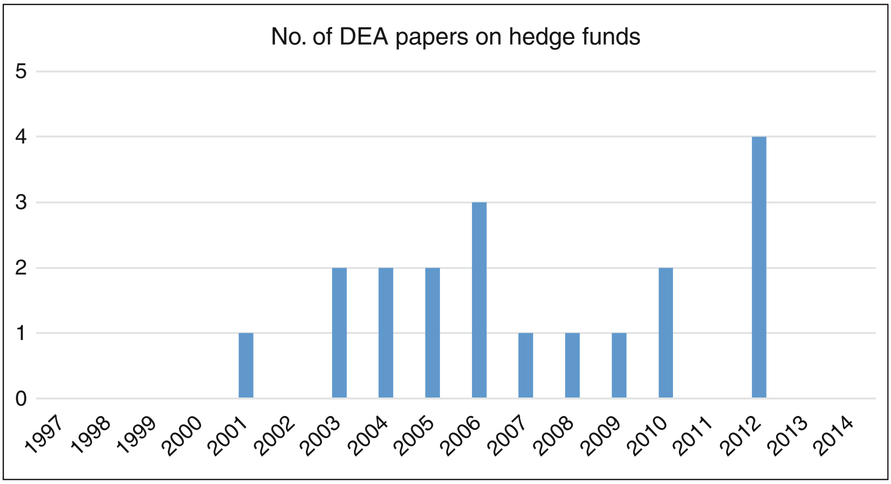


Fig. 8.2 Number of published papers on DEA models for the evaluation of the performance of hedge funds, CTAs and managed futures funds by year of publication

Furthermore, the review of the literature on DEA models for pension funds is presented in Table 8.4, while Table 8.5 considers the papers which try to apply DEA to exchange-traded funds (ETFs). With regard to the applications of DEA to ETFs, however, we must bear in mind that the aim of ETFs is to track a stock or a bond index, rather than to beat the market.

Another field of application of the DEA methodology in finance is the assessment of the performance of hedge funds. Table 8.6 displays the contributions of the DEA literature on hedge funds.

In addition, Table 8.7 reports the contributions that apply DEA to measure the performance of commodity trading advisors (CTAs) and managed futures funds.

On the other hand, mutual funds are actually stock and/or bond portfolios managed by financial professionals. Therefore, some DEA models that have been proposed for portfolio selection may also be useful for the performance evaluation of mutual funds. Table 8.8 presents the papers which use the DEA methodology for portfolio selection. As can be seen, the number of such papers has become considerable mainly in the last years; see also Fig. 8.3, that displays the number of published papers by year.

8.3 The Problem of Negative Returns in DEA Modeling

In classical DEA models it is common to assume that all the input and output values are non negative. This is indeed a crucial assumption in the measurement of performance with the DEA technique. However, it is far from being always satisfied

Table 8.1 Stylized review of the literature on DEA models for the assessment of the performance of mutual funds

Paper	DEA models	Features	Empirical analysis
Adeli O. (2013)	CCR	Uses various risk measures. Inputs: standard deviation, beta, half-variance, costs of issuance and cancellation. Outputs: downside, best time	Iran, 29 funds
Alexakis P., Tsolas I. (2011)	BCC	Empirical paper. Inputs: standard deviation, beta, assets, front-end load, back-end load. Output: return	Greece, 2001–2004, 55 funds
Ali Asghar M. J.-K., Afza T., Bodla M.A. (2013)		Empirical paper. Inputs: management fee, business services, equity. Outputs: invest-ments, returns	Pakistan, 2005–2010, 100 funds
Anderson, R.I., Brockman, C.M., Giannikos, C., McLeod, R.W. (2004)	CCR	Empirical analysis on real estate mutual funds. Inputs: standard deviation, front load, deferred load, 12b-1 fees, other expenses. Output: return	1997–2001, 110 real estate funds
Babalos V., Caporale G.M., Philippas N. (2012)	DEA-based Malmquist index	Evaluates the funds' total productivity change using the DEA-based Malmquist index. Employs a panel logit model to analyse the relationship between the probability of being efficient and funds' size. Inputs: standard deviation, total expense ratio, capital invested. Output: terminal value of the investment	Greece, 2003–2009, 31 funds
Babalos V., Doumpos M., Philippas N., Zopounidis C. (2015)	CCR, BCC	Combines DEA with MCDA. First uses DEA, then adopts a multicriteria decision aid (MCDA) approach to develop an overall per-formance measure on the basis of the DEA efficiency results. Inputs: gross expense ratio, turnover rate, standard deviation. Outputs: 1 +DMR (deviation from the median return), capital flow	US, 2003–2010, more than 500 funds

(continued)

Table 8.1 (continued)

Paper	DEA models	Features	Empirical analysis
Baghdadabada M.R.T., Tanha F.H., Halid N. (2013)		Evaluates funds managers' efficiency in terms of management style. Analyzes the benefits of the DEA approach in the DARA, CARA and IARA framework. Inputs: variance, turnover, expense ratio, loads. Outputs: return, stochastic dominance indicators	US, 2005–2010, 17,686 funds
Basso A., Funari S. (2001)	CCR	Proposes a single-output DEA performance measure that generalizes Sharpe, Treynor and reward-to-half-variance indices and a multiple output performance measure that takes into account a stochastic dominance indicator that reflects the investors' preference structure and the time occurrence of the returns. Inputs: some risk measures, subscription and redemption costs. Outputs: a return measure (expected return or expected excess return), stochastic dominance indicator	Italy, 1997–1999, 47 funds
Basso A., Funari S. (2005a)	CCR, cross-efficiency	Proposes a model that synthesizes the results of the traditional risk adjusted indices in a global performance measure which takes into account also the subscription and redemption costs and suggests an average cross-efficiency measure. Inputs: subscription and redemption costs, unit invested amount. Outputs: different traditional risk adjusted performance indices (Sharpe, reward to half-variance, Treynor, Jensen), stochastic dominance indicator	Italy, 1997–2001, 50 funds

Basso A., Funari S. (2014c)	BCC-O	Discusses the role of fund size in DEA performance evaluation and wonders whether it is appropriate to include size information among the input/output variables of DEA models. Moreover, analyzes the nature of returns to scale in mutual fund performance. Inputs: capital invested net of the initial fees, beta. Output: final value of the investment net of the exit fee	Europe, 2006–2009, 279 funds
Batra D., Batra G. (2012)		Compares systematic investment plan with lump sum investment using DEA. Inputs: standard deviation, beta, minimum investment. Outputs: 3 year return, 5 year return, Sharpe ratio	
Brandouy O., Kerstens K., Van de Woestyne I. (2015)	Shortage function	Explores the potential benefits of a series of non-parametric frontier-based fund rating models. Adopts a comparative approach based on a backtesting methodology	Europe, 2005–2011, 814 funds
Chang K.-P. (2004)	MCIRS	Uses a minimum convex input requirement set (MCIRS) approach to evaluate the performance of mutual funds. Inputs: beta, standard deviation, assets. Output: return	US, 1992–1996, 248 funds
Chang J.F. (2010)	CCR (DPEI model of Murthi, Choi, Desai, 1997)	Provides a method to create funds of funds, as well as strategies to choose sub-funds (with the DPEI index and a genetic algorithm) and a way to allocate weighted capital. Input and output variables are those of DPEI model of Murthi et al. (1997)	Taiwan, 2004–2006, 49 funds

(continued)

Table 8.1 (continued)

Paper	DEA models	Features	Empirical analysis
Chen Y.-C., Chiu Y.-H., Li M.-C. (2011)	System BCC model	Compares the results obtained with the BCC model and a system BCC model (with two subgroups of funds, stock and balanced) using as outputs the classical indicators of mutual fund performance. Inputs: standard deviation, purchasing turnover rate, direct transaction cost rate, selling expense rate. Outputs: Treynor index, Sharpe index, Jensen index, return rate	Taiwan, 2007, 278 funds
Chen Z., Lin R. (2006)	BCC-I	Considers several risk measures. Treats the same fund during different time periods as different funds. Inputs: investment costs, some risk measures, including value at risk and conditional value at risk. Outputs: a return measure (expected return or expected excess return), stochastic dominance indicator, traditional performance indices (Treynor index, Sharpe index, Jensen index)	China, 2000–2002, 33 funds; 1999–2000, 14 funds
Choi Y.K., Murthi B.P.S. (2001)	BCC	Compares DEA scores with the values obtained using the classical fund performance indices; analyzes the importance of fund's size and the returns to scale. Inputs: standard deviation, expense ratio, loads, turnover. Output: return	731 funds of different categories
Daraio C., Simar L. (2006)	DEA, FDH for comparison	Proposes a robust nonparametric approach for analysing mutual funds based on an order-m frontier and on a probabilistic approach; compares the results with those obtained with DEA and FDH. Inputs: standard deviation, expense ratio, turnover ratio. Output: return	US

de Guzman M.P.R., Cabanda E.C. (2005)	DEA-based Malmquist index	Empirical paper. Inputs: redemption costs, operating expenses. Outputs: total assets, number of accounts, total sales	Philippines, 1999–2003, 10 funds
Devaney M., Weber W.L. (2005)	Directional output distance function with CRS and VRS	Models risk as an undesirable output. Estimates from a directional output distance function are used to construct a risk/return frontier that defines the best-practice management technology for real estate investment trusts. Inputs: book value of assets, expenses. Desirable output: return. Undesirable output: risk (either variance or beta)	US, 1995–2000, 77 real estate investment trusts
Galagadera D.U.A., Silvapulle P. (2002)	CCR, BCC	Uses logistic regression to examine the dependence of the DEA efficiency scores on fund attributes, management strategy and operating environment. Seven inputs: four standard deviations of the 1, 2, 3, 5 year gross performance, sales charges, operating expenses, minimum initial investment. Four outputs to capture short, medium and long term performance	Australia, 1995–1999, 257 funds
Gökgöz F. (2010)	CCR, BCC	Empirical paper	Turkey, 2006–2007, 36 mutual and 41 pension funds
Gökgöz F., Çandarli D. (2011)	CCR, BCC-I	Compares technical, pure and scale efficiency of pension funds and mutual funds. Inputs: standard deviation, beta, expense ratios, turnovers. Output: excess return	Turkey, 2009, 36 mutual and 36 pension funds
Gregoriou G.N. (2006a)	CCR, super-efficiency, cross-efficiency	Empirical paper. Inputs: standard deviation, downside deviation, maximum drawdown. Outputs: return, monthly percentage profitable	US, 1990–2005, 25 funds
Gregoriou G.N. (2007)		Examines the efficiency of large US stocks, bonds and balanced funds using DEA. Compares the efficiency of the funds of the various DEA models with the Sharpe index	US

(continued)

Table 8.1 (continued)

Paper	DEA models	Features	Empirical analysis
Guo J., Ma C., Zhou Z. (2012)	BCC-I	Considers the higher order moment characteristics of funds return which reflect the preference of investors. Inputs: net assets, unit cost ratio, standard deviation, kurtosis. Outputs: accumulative daily net asset value, growth rate, skewness	China, 2010, 27 funds
Hanafizadeh P., Khedmatgozar H. R., Emrouznejad A., Derakhshan M. (2014)		Uses neural network back-propagation DEA in measurement of mutual funds efficiency	
Haslem J.A., Scheraga C.A. (2003)	BCC-I	Empirical paper. Inputs: cash, expense ratio, stocks, P/E ratio, P/B ratio, total asset. Output: Sharpe index	1999, 80 funds
Haslem J.A., Scheraga C.A. (2006)	BCC-I	Examines the efficiency of mutual fund portfolio management; follows Haslem, Scheraga (2003). Inputs: cash, expense ratio, stocks, P/E ratio, P/B ratio, number of securities held, portfolio turnover. Output: total asset	2001, 58 funds
Hsu C.-S., Lin J.-R. (2007)	CCR	Tests performance persistence. Inputs: standard deviation, ratio of management fee, ratio of load fee, turnover ratio. Output: excess return	Taiwan, 1999–2003, 192 funds
Hu J.-L., Yu H.-E., Wang Y.-T. (2012)	BCC-I	Uses an input oriented BCC model to adjust for a negative output; the truncated regression model is used to estimate effects of environmental variables on input slacks in a second stage. Inputs: standard deviation, expense ratio. Output: return	Taiwan, 2006–2010, 60 funds

Hu J.-L., Chang T.-P. (2008)	CCR	Applies a three-stage DEA approach to decompose mutual fund underperformance and to obtain pure managerial performance. Inputs: standard deviation, expense ratio. Output: return	Taiwan, 2005–2006, 156 funds
Joro T., Na P. (2006)		Proposes a non-linear DEA-like framework where the correlation structure among the units is taken into account and the diversification effects are calculated. Input: variance. Outputs: excess return, skewness	1995–2000, 54 funds
Kerstens K., Mounir A., Van de Woestyne I. (2011)	Shortage function	Argues in favor of the use of the shortage function and discusses three crucial specification issues in using non parametric methods for fund evaluation: (1) returns to scale issue (2) - higher-order moments and cost components issues (3) convexity issue	US, Europe, 2004–2009, 1070 funds
Khedmatgozar H.R., Kazemi A., Hanafizadeh P. (2013)	VEA, BCC-O	Uses value efficiency analysis (VEA) to incorporate the investor's judgements and compares the results with those of general DEA models. Input: variance. Outputs: return, skewness	US, 2005–2007, 22 funds
Lamb J.D., Tee K.-H. (2012a)	Input oriented NIRS, diversification-consistent DEA model	Discusses the assumptions needed for a DEA model of investment funds. Models diversification through a series of linear approximation to an ideal nonlinear model. Shows how the convex programme might be solved through an iterative procedure. Uses coherent risk measures	2000–2004, 30 hedge funds
Lamb J.D., Tee K.-H. (2012b)	Input and output oriented NIRS, diversification-consistent DEA model	Studies uncertainty in DEA efficiency estimates; investigates consistency and bias and uses bootstrap to develop stochastic DEA models for funds	2000–2004, 30 hedge funds

(continued)

Table 8.1 (continued)

Paper	DEA models	Features	Empirical analysis
Lin R. (2009)		Uses data envelopment windows analysis and considers the different period separately. Includes conditional value at risk as risk measure in the DEA model	
Lin R., Chen Z. (2008)		Uses value at risk and conditional value at risk as risk measures and treats the same fund during different time periods as different funds	US
Lozano S., Gutiérrez E. (2008a)	Additive metric	Proposes six DEA-like linear programming models consistent with second-order stochastic dominance, with different choices of input and output variables	Spain, 2002–2005, 108 funds
Lozano S., Gutiérrez E. (2008b)		Proposes three DEA-based models consistent with third-order stochastic dominance. Each model considers an appropriate risk measure as input and an appropriate return measure as output	Spain, 2002–2005, 33 funds
Majid M.S.A., Maulana H. (2010)	DEA-based Malmquist index	Empirical paper. Inputs: front-end loads, redemption fees, expense ratio; Output: return	Indonesia, 2004–2007, 23 mutual funds
Malhotra D.K., Malhotra R. (2012)	CCR	Empirical paper. Inputs: standard deviation, beta, load, 12b-1 charges, expense ratios. Output: return	2010–2011, 189 aggressive growth mutual funds
Margaritis D.M., Otten R., Tourani-Rad A. (2007)	BCC-I	Uses a Tobit regression to analyze the dependence of the efficiency scores on some fund attributes not included in the DEA analysis. Inputs: standard deviation, expenses ratio, load factor. Output: return	New Zealand, 1998–2003, 52 funds
Matallín-Sáez J.C., Soler-Domínguez A., Tortosa-Ausina E. (2014)	FDH, order-m, order-α partial frontiers	Applies DEA FDH, order-m and order-α partial frontiers. Inputs: standard deviation, kurtosis, expense ratio, beta. Outputs: return, skewness	US, 2001–2011, 1450 funds

McMullen P.R., Strong R.A. (1998)	BCC-I, BCC-O	Rescales the input and output variables so that they start from zero. Inputs: standard deviation, sales charge, minimum initial investment, expense ratio. Outputs: 1 year return, 3 year return, 5 year return	135 funds
Morey M.R., Morey R.C. (1999)	Quadratic programming	Presents two approaches, which take inspiration from data envelopment analysis, for identifying those funds that are strictly dominated. Inputs: variances over different horizons. Outputs: returns over different horizons	1985–1995, 26 funds
Murthi B.P.S., Choi Y.K., Desai P. (1997)	CCR	Proposes one of the first DEA models for the evaluation of mutual funds, the DEA portfolio efficiency index (DPEI). Inputs: standard deviation, expense ratio, loads, turnover. Output: return	731 funds of different categories
Pendaraki K. (2012)	BCC-I	Analyzes the effect of inclusion of higher moments as variables in DEA performance risk-return framework by applying DEA twice. i) Input: standard deviation. Outputs: return, asset. ii) Inputs: standard deviation, kurtosis. Outputs: return, asset, skewness	Greek, 2007–2010, 43 funds
Premachandra I.M., Zhu, J., Watson J., Galagadera D.U.A. (2012)	Two-stage DEA	Uses two-stage DEA models that decomposes the overall efficiency into two components: operational efficiency and portfolio efficiency. Stage 1 has two inputs (management fees, 12b-1 fees) and one output (net asset value). Stage 2 has five inputs (fund size, net expense ratio, turnover, standard deviation, adjusted net asset value) and one output (return)	US, 1993–2008, 66 families (1269 mutual funds)

(continued)

Table 8.1 (continued)

Paper	DEA models	Features	Empirical analysis
Simar L., Vanhems A., Wilson P.W. (2012)	FDH and DEA estimators, directional distance function, VRS	Studies statistical properties of directional estimators; establishes that the directional DEA estimators share the properties of the traditional radial DEA estimators. Inputs: standard deviation, expense ratio, turnover ratio. Output: return	2001, 129 aggressive-growth funds
Soongswang A., Sanohdontree Y. (2011a)	CCR	Empirical paper	Thailand, 2002–2007, 138 funds
Soongswang A., Sanohdontree Y. (2011b)		Empirical paper	Thailand, 2002–2007, 138 funds
Tarim S.A., Karan M.B. (2001)	CCR with restrictions on weights	Uses the DPEI approach of Murthi et al. (1997) with the introduction of upper and lower bounds on the weights of the input variables. Inputs: standard deviation, expense ratio, turnover ratio. Output: return	Turkey, 1998, 191 funds
Tavakoli Baghdadabad M.R., Noori Houshyar A. (2014)	CCR, DEA-based Tornqvist index	Assesses the changes of mutual funds' total productivity using the DEA-based Tornqvist productivity index. Moreover, uses a panel logit model to study the relationship between efficiency and productivity and some funds' features	US, 2000–2012, 11,522 funds
Tsolas I.E. (2014)	BCC-I	Uses a two-stage procedure with first a BCC model and then a Tobit model. Inputs: standard deviation, management expense ratio, front load, deferred load. Output: return	62 precious metal funds
Zhao X., Shi J. (2010)		Uses a coned context DEA model with expected shortfall modeled under an asymmetric Laplace distribution. Input: expected shortfall. Outputs: 1 year return, 3 year return, 5 year return	China, 2003–2007, 17 funds

Zhao X.-J., Wang S.-Y. (2007)	CCR, BCC-O	Empirical paper. Inputs: comparative standard deviation, percentage of negative monthly return, operational fees. Outputs: final value, return	China, 2004–2005, 78 funds (24 open-end and 54 closed-end funds)
Zhao X., Wang S., Lai K.K. (2011)		Proposes two quadratic-constrained DEA models to evaluate mutual funds, based on endogenous benchmarks	China, 2005–2006, 25 funds
Zhao X., Yue W. (2008)		Proposes a coned context dependent DEA model	China, funds
Zhao X., Yue W. (2012)	MFDEA	Proposes a multi-subsystem fuzzy data envelopment analysis (MFDEA) model to evaluate mutual funds management companies' core competence. Considers two subsystems with different choices of input and output variables	China, 2004–2008, 32 mutual funds management companies

Table 8.2 Stylized review of the literature on DEA models for the assessment of the performance of socially responsible investment (SRI) mutual funds

Paper	DEA models	Features	Empirical analysis
Abdelsalam O., Duygun-Fethi M., Matallín J.C., Tortosa-Ausina E. (2014a)	Two variants of FDH	Combines two variants of FDH methods (order-m and order-alpha) in the first stage of analysis and quantile regression in the second stage. Inputs: standard deviation, kurtosis, expenses. Outputs: gross return, skewness	With a geographic focus (East/West/Middle/global), 2001–2011, 138 Islamic and 636 SRI funds
Abdelsalam O., Duygun M., Matallín-Sáez J.C., Tortosa-Ausina E. (2014b)	Two variants of FDH	Uses two efficiency models based on FDH (order-m and order-alpha) in the first stage of analysis, a recursive investment approach in the second stage; persistence performance is evaluated in the third stage. Inputs: standard deviation, kurtosis, expenses. Outputs: gross return, skewness	With a geographic focus (East/West/Middle/global), 2001–2011, 138 Islamic and 636 SRI funds
Basso A., Funari S. (2003)	CCR, exogenously fixed output, categorical variable model	Proposes three different DEA models for the performance evaluation of SRI mutual funds: a basic model, an exogenously fixed output model and a categorical model. Inputs: standard deviation, beta, subscription costs, redemption costs. Outputs: expected return, ethical indicator	Simulated data
Basso A., Funari S. (2005b)	Exogenously fixed output	Proposes a DEA model for the performance evaluation of ethical funds with non negative outputs and an exogenously fixed ethical level. Inputs: initial investment, standard deviation, initial fees, exit fees. Outputs: final value of the investment, ethical level	UK, 2002–2005, 32 SRI and 25 - non-SRI funds
Basso A., Funari S. (2008)	CCR, exogenously fixed output	Proposes a measure of the ethical level of SRI funds which takes into account the main socially responsible features and three DEA models that adjust the ones proposed in Basso	Europe, 2002–2005, 159 SRI and 110 - non-SRI funds

		and Funari (2003) to tackle the problem of the presence of negative returns. Inputs: initial investment, standard deviation, initial fees, exit fees. Outputs: final value of the investment, ethical level	
Basso A., Funari S. (2014a)	CCR, BCC-O, exogenously fixed output	Proposes some models which can be computed in all phases of the business cycle. Tests the presence of returns to scale on European SRI equity funds; provides a measure of the degree of social responsibility of SRI funds; compares the performances between SRI and non-SRI funds. Inputs: initial payout invested in the fund, beta. Outputs: final value of the investment, ethical level	Europe, 2006–2009, 189 SRI and 90 - non-SRI funds
Basso A., Funari S. (2014b)	Models with a constant input; CCR, BCC-O, exogenously fixed output	Proposes constant and variable returns to scale DEA models which consider a constant initial capital and standard deviation as inputs and the final value of the investment and ethical level as outputs. The implications of the presence of a constant input in DEA models are studied	Europe, with special focus on Sweden, 2006–2009, 189 SRI and 90 non-SRI funds
Basso A., Funari S. (2014d)	CCR, exogenously fixed output	Evaluates the performance of SRI equity funds in the main European countries with three different DEA models, compares the performance of SRI and non-SRI mutual funds in the various countries with a series of statistical tests and compares the performance obtained by SRI mutual funds among different countries. Input and output variables are those considered in Basso and Funari (2008)	Europe, 2006–2009, 190 SRI and 91 - non-SRI funds

(continued)

Table 8.2 (continued)

Paper	DEA models	Features	Empirical analysis
Einolf K.W. (2007)	CCR	Uses DEA in an application of SRI portfolio development to find the best financially and socially performing companies within each industry sector (with a best-in-class optimization approach in place of screening). Inputs: IW financial environment, social and governance scores. Outputs: alpha, Morningstar star rating	US, 2007, 978 stocks
Jeong S.O., Hoss A., Park C., Kang K.-H., Ryu Y. (2013)	DEA, FDH; input oriented approach	Provides a stock portfolio screening tool for socially responsible investment based upon corporate social responsibility and financial performance. Compares DEA and FDH results. Inputs: inverted ESG (environmental, social, governance) scores. Outputs: return on assets, return on equity, operating profit percentage	Korea, 2006–2009, 253 stocks
Pérez-Gladish B., Méndez Rodríguez P., M'zali B., Lang P. (2013)	Output oriented VRS	Integrates the role of financial criteria and the influence of social responsibility considerations in investment decisions within a DEA framework consistent with second-order stochastic dominance. Compares the performance of conventional and SRI funds. Inputs: conditional value at risk, turnover ratio, expense ratio, deferred loads, front loads. Outputs: return, social and environmental responsibility index, quality of socially responsible investment management	US, 2007, 25 SRI and 21 conventional funds

Table 8.3 Stylized review of the literature on DEA models for the assessment of the performance of Islamic mutual funds

Paper	DEA models	Features	Empirical analysis
Abdelsalam O., Duygun-Fethi M., Matallín J.C., Tortosa-Ausinad E. (2014a)	Two variants of FDH	Combines two variants of FDH methods (order-m and order-alpha) in the first stage of analysis and quantile regression in the second stage. Inputs: standard deviation, kurtosis, expenses. Outputs: gross return, skewness	With a geographic focus (East/West/ Middle/global), 2001–2011, 138 Islamic and 636 SRI funds
Abdelsalam O., Duygun M., Matallín-Sáez J.C., Tortosa-Ausina E. (2014b)	Two variants of FDH	Uses two efficiency models based on FDH (order-m and order-alpha) in the first stage of analysis, a recursive investment approach in the second stage; persistence performance is evaluated in the third stage. Inputs: standard deviation, kurtosis, expenses. Outputs: gross return, skewness	With a geographic focus (East/West/ Middle/global), 2001–2011, 138 Islamic and 636 SRI funds
Majid M.S.A., Maulana H. (2012)		Investigates the influence of the mutual funds companies' characteristics on efficiency measures using DEA and the generalized least square estimation. Compares the performance of Islamic and conventional mutual funds companies	Indonesia, 2004–2007
Rubio J.F., Hassan M.K., Merdad H.J. (2012)	BCC-I, Russell model	Performance comparison of Islamic and non Islamic mutual funds. Inputs: standard deviation, lower partial momentum, maximum drawdown period. Outputs: expected return, upper partial momentum, maximum period of consecutive gain	Indonesia
Saad N.M., Majid M.S.A., Kassim S., Hamid Z., Yusof, R.M. (2010)	DEA-based Malmquist index	Investigates the efficiency of conventional and Islamic unit trust companies. Inputs: management expenses ratio, portfolio turnover ratio. Output: return	Malaysia, 2002–2005, 5 Islamic and 22 unit trust companies

Table 8.4 Stylized review of the literature on DEA models for the assessment of the performance of pension funds

Paper	DEA models	Features	Empirical analysis
Andreu L., Sarto J.L., Vicente L. (2013)	SBM	Applies four variants of the slacks-based measure (SBM) to evaluate the efficiency of strategic asset allocation in pension fund management	
Barrientos A., Boussofiane A. (2005)	CCR, BCC	Studies the performance of pension fund management market over time using DEA. Inputs: marketing and sales costs, office personnel and executive pay, administration and computing costs. Outputs: total revenue, number of contributors	Chile, 1982–1999, pension funds
Gökgöz F. (2010)	CCR, BCC	Empirical paper	Turkey, 2006–2007, 36 mutual and 41 pension funds
Gökgöz F., Çandarli D. (2011)	CCR, BCC-I	Compares technical, pure and scale efficiency of pension funds and mutual funds. Inputs: standard deviation, beta, expense ratios, turnovers. Output: excess return	Turkey, 2009, 36 mutual and 36 pension funds
Guillén G.B. (2008)	CCR	Uses DEA to compare pension fund institutions of nine Latin American countries. Inputs: administrative cost, sale cost. Outputs: total revenue, number of contributors	Latin America, 2005–2007, 876 pension fund institutions of 9 countries
Medeiros Garcia M.T. (2010)	DEA-based Malmquist index	Estimates the change in total productivity, the technically efficient change and the technological change by means of DEA-Malmquist index. Inputs: value of pensions paid, number of workers, net assets, contributions received. Outputs: number of funds managed, value of the funds managed, profits paid to stockholders, number of participants	Portugal, 1994–2007, 12 pension funds management companies
Pestana Barros C., Medeiros Garcia M.T. (2006)	Cross-efficiency, super-efficiency	Uses DEA to evaluate the performance of pension funds management companies with the cross-efficiency and the super-efficiency DEA model. Inputs: number of full time equivalent workers, fixed assets, contributions received from participants or sponsors. Outputs: number of funds managed, value of the funds managed, pensions paid to subscribers	Portugal, 1994–2003, 12 pension funds management companies

Table 8.5 Stylized review of the literature on DEA models for exchange-traded funds (ETFs)

Paper	DEA models	Features	Empirical analysis
Chu J., Chen F., Leung P. (2010)	RDM	Applies DEA to evaluate the performance of IShare World exchange-traded funds with a range directional measure (RDM)	2006–2009, IShare World ETFs
Prasanna P.K. (2012)	Cross-efficiency, super-efficiency	Uses cross-efficiency and super-efficiency DEA models to evaluate the performance of ETFs. Inputs: standard deviation, maximum drawdown, downside deviation. Outputs: percentage of months with positive returns, compounded monthly return	India, 2006–2011, 82 ETFs
Tsolas I.E. (2011)	Generalized proportional distance function	Applies a two-stage procedure. In the first stage, the generalized proportional distance function in the DEA context is used to measure the relative efficiency of ETFs. In the second stage, a Tobit model is employed to identify the drivers of performance. Inputs: portfolio price/cash flow, portfolio price/book, total expense ratio. Output (first DEA run): Sharpe index. Output (second DEA run): Jensen's alpha. Output (third DEA run): Sharpe ratio, Jensen's alpha	2008–2010, 15 natural resources ETFs
Tsolas I.E., Charles V. (2015)	RAM	Appraises the performance of a sample of ETFs using the range-adjusted measure (RAM), the RAM-BCC model and a common set of weights of RAM	

Table 8.6 Stylized review of the literature on DEA models for the assessment of the performance of hedge funds

Paper	DEA models	Features	Empirical analysis
Ammann M., Moerth P. (2008)		Discusses the impact of fund size on the performance of funds of hedge funds. Inputs: standard deviation, drawdown, kurtosis, modified value at risk. Outputs: return, skewness, proportion of positive months, omega, Sortino ratio, kappa, upside potential ratio, Calmar ratio, alpha	2000–2005, 167 funds of hedge funds
Eling M. (2006)	CCR, BCC, super-efficiency	Presents ten DEA applications with different choices of input and output variables. In some applications the choice of inputs and outputs is based on two selection rules (Spearman's rank correlation, principal component analysis)	1996–2005, 30 hedge funds
Gregoriou G.N. (2003)	BCC, cross-efficiency, super-efficiency	Uses three DEA models to evaluate the management quality of funds of hedge funds. Inputs: lower mean semi-skewness, lower mean semi-variance, mean lower return. Outputs: upper mean semi-skewness, upper mean semi-variance, mean upper return	1997–2001, 168 funds of hedge funds
Gregoriou G.N., Rouah F., Sedzro K. (2003)		Chapter 4: Fund of hedge fund performance using data envelopment analysis	
Gregoriou G.N., Sedzro K., Zhu J. (2005)	BCC, cross-efficiency, super-efficiency	Uses three DEA models to obtain insights due to problems encountered when using multifactor models to predict hedge fund returns. Inputs: lower mean semi-skewness, lower mean semi-variance, mean lower return. Output: upper mean semi-skewness, upper mean semi-variance, mean upper return	1997–2001, 614 hedge funds

Gregoriou G.N., Zhu, J. (2005)	CRS, VRS, context-dependent DEA, fixed- and variable-benchmark models	Book on DEA performance of hedge funds and CTAs. Different types of models with different input and output variables	1998–2004, top 20 hedge funds, funds of hedge funds and CTAs from various categories
Gregoriou G.N., Zhu J. (2007)	BCC-I	Applies DEA to funds of hedge funds. Inputs: standard deviation, downside deviation, maximum drawdown. Outputs: percentage of profitable months, return, number of successive months with positive returns	1994–2004, 25 funds of hedge funds
Kumar U.D., Roy A., Saranga H., Singal K. (2010)	SBM, super-SBM	Uses slack-based measure (SBM) and super-SBM models. Analyzes the hedge fund strategies using a variety of classical risk return measures. Based on Spearman's rank correlation the following inputs and outputs are considered: standard deviation of drawdown and value at risk as inputs, higher partial moment of order 0, skewness as outputs. Moreover, a principal component analysis is carried out	1995–2007, 4730 hedge funds
Lamb J.D., Tee K.-H. (2012a)	Input oriented NIRS, diversification-consistent DEA model	Discusses the assumptions needed for a DEA model of investment funds. Models diversification through a series of linear approximation to an ideal nonlinear model. Shows how the convex programme might be solved through an iterative procedure. Uses coherent risk measures	2000–2004, 30 hedge funds
Lamb J.D., Tee K.-H. (2012b)	Input and output oriented NIRS, diversification-consistent DEA model	Studies uncertainty in DEA efficiency estimates; investigates consistency and bias and uses bootstrap to develop stochastic DEA models for funds	2000–2004, 30 hedge funds
Lamb J., Tee K.-H. (2012c)	Input oriented NIRS, diversification-consistent DEA model	Discusses about hedge funds risk. Uses the diversification-consistent DEA model of Lamb and Tee (2012a)	2000–2004, 30 hedge funds
Sedzro K. (2009)	BCC	Compares the rankings obtained with the stochastic dominance method and DEA	1998–2003; 615 hedge funds

Table 8.7 Stylized review of the literature on DEA models for the assessment of the performance of commodity trading advisors (CTAs) and managed futures funds

Paper	DEA models	Features	Empirical analysis
Darling G., Mukherjee K., Wilkens K. (2004)	BCC-I	Investigates the performance of CTAs with a DEA model and Tobit regressions of the scores on various features. Inputs: standard deviation, proportion of negative returns. Outputs: return, minimum return, skewness, kurtosis	1998–2002, 216 CTAs
Diz F., Gregoriou G.N., Rouah F., Satchell S.E. (2004)	BCC-I, cross-efficiency, super-efficiency	Studies the efficiency of CTAs using DEA; computes also cross-efficiencies and super-efficiencies. Inputs: lower mean semi-skewness, lower mean semi-variance, mean lower return. Outputs: upper mean semi-skewness, upper mean semi-variance, mean upper return	1997–2001, 150 CTAs
Glawischnig M., Sommersguter-Reichmann M. (2010)	BCC-I, assurance region model, Russell measure, RAM	Compares various existing DEA models for the evaluation of the performance of alternative investment fund industry. Inputs: lower partial moments 0–3, maximum drawdown periods. Outputs: upper partial moments 1–3	2004–2007, 167 managed futures funds
Gregoriou G.N. (2006b)	BCC-I, cross-efficiency, super-efficiency	Computes DEA efficiency, cross-efficiency and super-efficiency of CTAs. Inputs: beginning assets under management, number of futures contracts traded per million dollars net of all brokerage fees. Outputs: compound return, ending assets under management	1995–2004, 90 CTAs
Gregoriou G.N., Chen Y. (2006)	Input oriented VRS	Investigates the performance of CTAs using fixed and variable benchmarking models. Inputs: standard deviation, semi-deviation, percentage of negative returns, periods to recover from first largest maximum drawdown. Outputs: compounded return, percentage of positive returns	1998–2004, 143 CTAs

Gregoriou G.N., Henry S.C. (2015)	BCC-O with undesirable outputs	Investigates the efficiency of CTAs using undesirable outputs. Inputs: margin to equity ratio, incentive fee, round turn, management fee. Outputs: Sharpe index, cumulative return, maximum drawdown, standard deviation, semi-deviation	2009–2013, 50 CTAs
Gregoriou G.N., Zhu, J. (2005)	CRS, VRS, context-dependent DEA, fixed- and variable-benchmark models	Book on DEA performance of hedge funds and CTAs. Different types of models with different input and output variables	1998–2004, top 20 hedge funds, funds of hedge funds and CTAs from various categories
Tokic D. (2012)	BCC-O	Uses DEA to evaluate the relative performance efficiency of CTAs	1992–2010, 30 CTAs
Wilkens K., Zhu J. (2001)	BCC-I	Applies DEA to commodity trading advisor returns, using the input oriented BCC to deal with negative outputs. Inputs: standard deviation, proportion of negative returns. Outputs: average return, minimum return, skewness	1999, 11 CTAs

Table 8.8 Stylized review of the literature on DEA models for portfolio selection and stock indices

Paper	DEA models	Features	Empirical analysis
Adamauskas S., Krusinskas R. (2012)	CCR	Proposes a model to evaluate the efficiency of private investor's decisions with five stages; the fourth stage, regarding the investment instruments selection, uses DEA. Considers different input and output variables for stocks and funds evaluation	Lithuania, 2004–2010, 18 investment portfolios
Bahrani R., Khedri N. (2013)	CCR, BCC-I, BCC-O	Creates a portfolio of efficient companies using DEA. Inputs: assets, salaries, sale costs. Outputs: income, operating profit	Iran, 2001–2007, companies
Banihashemi S., Sanei M. (2013)		Evaluates the performance of portfolios and asset allocations using DEA and DEA/AHP (Analytic Hierarchy Process) ranking model. Input: variance. Output: return	
Branda M. (2011)	Quadratic programming	Proposes models related to the third-degree stochastic dominance (TSD), based on necessary conditions for TSD and on related mean-risk models; these models draw their inspiration from the DEA methodology and lead to quadratic programming problems	World, 2006–2010, 25 financial indices
Branda M. (2013a)	Mixed-integer linear programming	Proposes efficiency models which draw their inspiration from the DEA methodology and take into account portfolio diversification; they lead to mixed-integer linear programming problems. Uses general deviation measures as inputs and return measures as outputs	World, 2006–2010, 25 financial indices
Branda M. (2013b)		Presents numerically tractable formulations of the diversification-consistent models proposed in others contributions of Branda	US, 2002–2011, 46 industries portfolios
Branda M. (2013c)	Same model used in Branda (2013a)	Presents a proof of equivalence	
Branda M., Kopa M. (2012a)	CRS	Compares DEA, mean-risk and stochastic dominance on a set of financial indices. Inputs: several risk measures. Output: return	World, 2006–2010, 25 financial indices

Branda M., Kopa M. (2012b)	CRS, VRS	Compares several DEA models for portfolio efficiency and a diversification-consistent model. Inputs: several risk measures. Output: return	US, 1982–2011, 48 industries portfolios
Branda M., Kopa M. (2014)	CRS, VRS	Empirically compares CRS and VRS DEA models and diversification-consistent models. Inputs: conditional value at risk at several probability levels. Output: return	US, 1982–2011, 48 industries portfolios
Briec W., Kerstens K., Jokung O. (2007)	Shortage function	Extends the shortage function to the mean-variance-skewness space to account for a preference for positive skewness in addition to a preference for returns and an aversion to risk	1997–1999, 35 assets
Briec W., Kerstens K., Lesourd J.B. (2004)	Shortage function	Studies existing nonparametric efficiency measurement approaches for single period portfolio selection from a theoretical perspective and generalizes currently used efficiency measures into the mean-variance space	26 funds analyzed in Morey and Morey (1999)
Chen H.-H. (2008)		Uses DEA to construct stock portfolios	Taiwan, stocks
Dia M. (2009)		Presents a four-step methodology for portfolio selection based on DEA	Stocks
Ding H. Zhou Z., Xiao H., Ma C., Liu W. (2014)	BCC-I, BCC-O	Develops DEA models to evaluate the performance of portfolios with margin requirements; the BCC frontiers approximate the exact frontier. Input: variance. Output: expected return	Simulations
Edirisinghe N.C.P., Zhang X. (2007)	GDEA	Proposes a generalized DEA approach (GDEA) for stock portfolios in which the selection of inputs and outputs is sought iteratively. Potential inputs and outputs: 18 financial parameters through a range of performance perspectives (profitability, asset utilization, liquidity, leverage, valuation, growth)	US, 1996–2002, 230 stocks
Edirisinghe N.C.P., Zhang X. (2010)	Iterative two-stage optimization	Presents a methodology for selecting input and output variables endogenously to the DEA model in the presence of expert's knowledge	US, 1997–2005, 827 stocks

(continued)

Table 8.8 (continued)

Paper	DEA models	Features	Empirical analysis
Galagedera D.U.A. (2013)	CCR, cross-efficiency	Estimates the cross-efficiency of equity markets in a multi-dimensional risk-adjusted return framework. Inputs: standard deviation, beta, downside deviation. Outputs: two return factors (two positive variables based on observed excess returns)	World, 2003–2011, 40 equity markets
Gnanasekar I.F., Arul R. (2013)		Case study on financial risk tolerance using DEA	
Gnanasekar I.F., Arul R. (2014)		Studies the efficiency of portfolio investors with respect to the financial risk tolerance using DEA	
Hsu C.-M. (2014)	CCR	Proposes an integrated procedure using DEA, ant colony optimization for continuous domains and gene expression programming, in which DEA is used to select stocks in the first stage. Inputs: total assets, total equity, cost of sales, operating expenses. Outputs: net sales, net income	Taiwan, 2007–2011, 48 stocks
Huang C.-Y., Chiou C.-C., Wu T.-H., Yang S.-C. (2015)		Proposes an integrated DEA-MODM method for portfolio optimization. Uses DEA to select the portfolio and develops a multi-objective decision-making (MODM) model to determine the allocation of capital to each stock in the constructed portfolio	Taiwan
Huang T.H., Leu Y.H. (2014)		Presents a method to construct a profitable portfolio of mutual funds. In the first stage, the DEA, Sharpe and Treynor indices and the monthly rates of return are used to select a mutual fund portfolio. In the second stage, the linear regression model, the fruit fly optimization algorithm and the general regression neural network are used to construct a prediction model	
Ismail M.K.A., Salamudin N., Rahman N. M.N.A., Kamaruddin B.H. (2012)		Employs DEA to evaluate the firms' efficiencies, which are then used to form a portfolio	Malaysia, 2004–2005

Kadoya S., Kuroko T., Namatame T. (2008)	DEA, inverted DEA	Proposes an index for an investment strategy to capture the return-reversal effect using both DEA and inverted DEA. DEA inputs (inverted DEA outputs): three surprise indices calculated using sales, operating profit, ordinary profit. DEA outputs (inverted DEA inputs): 1 year return, 3 year return, 5 year return	Japan, 2000–2004, 1146 stocks
Kumar Singh A., Sahu R., Bharadwaj S. (2010)	CCR	Compares ordered weighted averaging (OWA)-heuristic algorithm and basic DEA for asset selection. Input: variance. Output: return	India, 2005–2007, 45 stocks
Lim S., Oh K.W., Zhu, J. (2014)		Uses DEA cross-efficiency evaluation in portfolio selection. Sixteen financial metrics within various performance perspectives are used as input and output variables. Input perspectives: asset utilization, liquidity, leverage. Output perspectives: profitability, growth	Korea, 2002–2011, stocks
Lopes A., Lanzer E., Lima M., da Costa N. Jr. (2008)	CCR	Defines a multi-period investment strategy based on DEA to select efficient stocks. Inputs: price to earnings ratio, beta, volatility. Outputs: 1 year return, 3 year return, 5 year return	Brazil, 2001–2006, stocks
Pai V. (2012)		Explores the application of an extended ant colony optimization algorithm for the solution of a risk budgeted portfolio optimization problem. Compares the results with those obtained by two other metaheuristic optimization methods. Uses DEA to compare the efficiencies of the optimal risk budgeted portfolios obtained by the three approaches	Bombay and Tokyo Stock Exchange, 2001–2006
Pätäri E., Leivo T., Honkapuro S. (2010)		Uses DEA as a basis of value portfolio selection criterion. The performance of portfolios is evaluated on the basis of average return and several risk-adjusted performance metrics	Finland, non-financial stocks

(continued)

Table 8.8 (continued)

Paper	DEA models	Features	Empirical analysis
Pätäri E., Leivo T., Honkapuro S. (2012)	CCR, super-efficiency, cross-efficiency	Uses DEA as a basis of selection criteria for equity portfolios, integrating the benefits of both value investing and momentum investing. Considers three variants of DEA models and four combinations of input and output variables	Finland, 1994–2010, 126 - non-financial stocks
Powers J., McMullen P.R. (2000)		Uses DEA to determine efficient stocks. Inputs: standard deviation, beta, price to earnings ratio. Outputs: 1 year return, 3 year return, 5 year return, 10 year return, earnings per share	185 large cap stocks
Premachandra I., Powell J.G., Shi J. (1998)	SDEA	Uses a spreadsheet-based numerical stochastic data envelopment analysis (SDEA) model estimated with a simulation package to study the selection of portfolios by fund managers. Inputs: total value initially invested in risky holdings, value of each portfolio's initial risk-free investments. Output: portfolio's total market value at the end of a time period minus the comparison benchmark return	New Zealand, 1975–1992
Sahoo B.K., Meera E. (2008)	SBM, CRS, VRS	Empirical paper on large cap market securities which uses different DEA models. Inputs: standard deviation, beta, price to earnings ratio. Outputs: 1 year return, 3 year return, 6 year return, earning per share	India, 93 stocks
Sengupta J.K. (2003)	BCC	Develops a set of nonparametric tests which includes the convex hull method and the stochastic dominance criteria. Different selections of input and output variables	1988–1998, 60 fund portfolios
Zamani L., Beegam R., Borzoian S. (2014)	BCC-I, super-efficiency	Uses DEA to build portfolios. Inputs: beta, modified 5 year beta, debt to equity ratio. Outputs: return on equity, return on capital employment, net profit margin, earning per share	India, 2013, 43 stocks

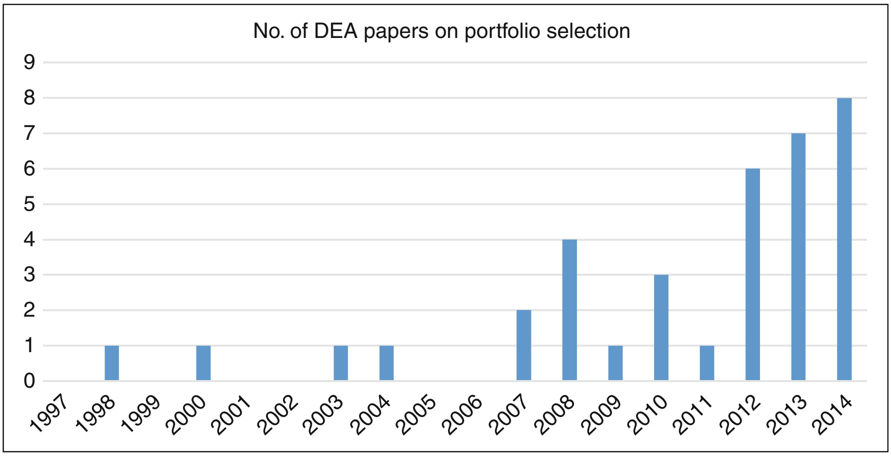


Fig. 8.3 Number of published papers on DEA models for portfolio selection and stock indices by year of publication

in the empirical applications to the performance evaluation of mutual funds when, as is usual, an output variable is chosen as either the mean return or the mean excess return.

On the other hand, it is well known in the literature that the DEA performance measure may give non satisfactory results when some output variables may take negative values; see for example Silva Portela et al. (2004). The reason can easily be grasped from a quick analysis of the following stylized examples, in which we consider a DEA model with constant returns to scale.

Let us consider the problem of evaluating the performance of four decision making units (DMUs) U_1, U_2, U_3, U_4, with one input x and two outputs y_1 and y_2. Let the values of the outputs of the four DMUs, normalized with respect to the input value, be as follows:

$$U_1 = \left(\frac{y_{11}}{x_1}, \frac{y_{21}}{x_1}\right) = (5, 1) \tag{8.1}$$

$$U_2 = \left(\frac{y_{12}}{x_2}, \frac{y_{22}}{x_2}\right) = (3, 2) \tag{8.2}$$

$$U_3 = \left(\frac{y_{13}}{x_3}, \frac{y_{23}}{x_3}\right) = (2, 3) \tag{8.3}$$

$$U_4 = \left(\frac{y_{14}}{x_4}, \frac{y_{24}}{x_4}\right) = (-1, a), \tag{8.4}$$

with $a \in \mathbf{R}^+$.

In the CCR model with constant returns to scale, the DEA performance measure for DMU j_0, with $j_0 \in \{1, 2, 3, 4\}$, is the optimal value of the following linear fractional programming problem

$$\max_{v, u_1, u_2} \frac{u_1 y_{1j_0} + u_2 y_{2j_0}}{v x_{j_0}} \tag{8.5}$$

s.t.

$$\frac{u_1 y_{1j} + u_2 y_{2j}}{v x_j} \leq 1 \quad j = 1, 2, 3, 4 \tag{8.6}$$

$$v, u_1, u_2 \geq \varepsilon, \tag{8.7}$$

where v, u_1, u_2 are the weights associated to the input and output variables, respectively, and ε is a non-Archimedean constant. The optimal solution can be found by solving the following equivalent output oriented linear program

$$\min_{v, u_1, u_2} v x_{j_0} \tag{8.8}$$

s.t.

$$u_1 y_{1j_0} + u_2 y_{2j_0} = 1 \tag{8.9}$$

$$u_1 y_{1j} + u_2 y_{2j} \leq v x_j \quad j = 1, 2, 3, 4 \tag{8.10}$$

$$v, u_1, u_2 \geq \varepsilon, \tag{8.11}$$

which is equivalent to the following reduced linear problem

$$\min_{v, u_1, u_2} v x_{j_0} \tag{8.12}$$

s.t.

$$u_1 \frac{y_{1j_0}}{x_{j_0}} + u_2 \frac{y_{2j_0}}{x_{j_0}} = \frac{1}{x_{j_0}} \tag{8.13}$$

$$u_1 \frac{y_{1j}}{x_j} + u_2 \frac{y_{2j}}{x_j} \leq v \quad j = 1, 2, 3, 4 \tag{8.14}$$

$$u_1, u_2 \geq \varepsilon. \tag{8.15}$$

If we restrict the analysis to the set of DMUs U_1, U_2, U_3, we have a classical DEA problem in which all the input and output values are positive. The efficiency frontier of such an instance is represented in Fig. 8.4, where the cartesian axes represent the normalized output values $\frac{y_{1j}}{x_j}$ and $\frac{y_{2j}}{x_j}$.

The efficient frontier is the upper-right line which connects the efficient DMUs, i.e. the DMUs with a DEA performance measure equal to 1. Figure 8.4 shows that DMUs U_1 and U_3 are efficient, while U_2 is inefficient and its DEA performance measure is equal to $\frac{dist\ (O, U_2)}{dist\ (O, P_2)} = 0.923$. As is well known, the point P_2 represents

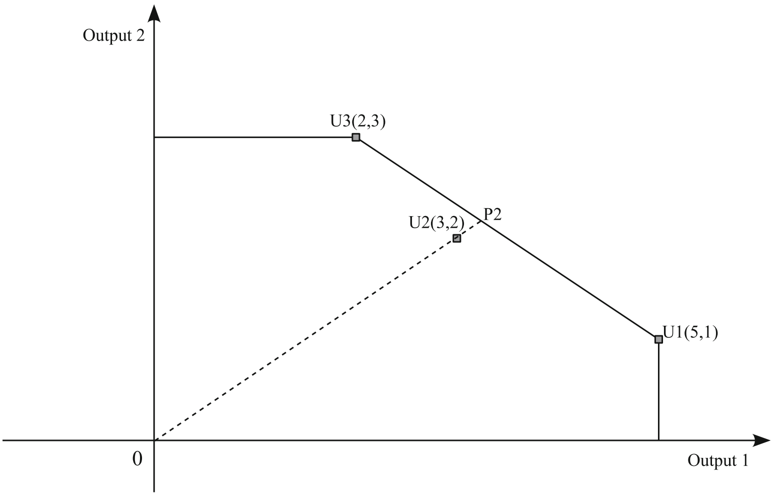


Fig. 8.4 Efficient frontier of the example with DMUs U_1, U_2, U_3 (normalized output values)

Table 8.9 DEA efficiency scores for DMUs U_1, U_2, U_3, U_4 for different values of $a \in \mathbf{R}^+$, the second output of U_4

a	Score of U_1	Score of U_2	Score of U_3	Score of U_4
1	1.000	0.923	1.000	0.333
2	1.000	0.923	1.000	0.667
3	1.000	0.923	1.000	1.000
4	1.000	0.923	1.000	1.000
5	1.000	0.923	1.000	1.000
6	1.000	0.871	0.903	1.000
7	1.000	0.833	0.833	1.000
8	1.000	0.805	0.780	1.000
9	1.000	0.783	0.739	1.000
10	1.000	0.765	0.706	1.000

the virtual unit which has the same input and output orientation as U_2 and lies on the efficient frontier. This virtual unit suggests that unit U_2 might improve its output values while keeping the input value fixed, by moving along the dashed line $\overline{OP_2}$ towards the efficient frontier, till its reaches efficiency.

If we include in the analysis also DMU U_4, which has a negative value of output 1, puzzling results can be obtained, so that the DEA fractional problem (8.5)–(8.7) does not give a reasonable efficiency measure any longer. Table 8.9 displays the values of the efficiency measure obtained for the four DMUs for different values of the second output of U_4, a; Figs. 8.5, 8.6, 8.7, and 8.8 show the efficient frontier obtained in some relevant cases.

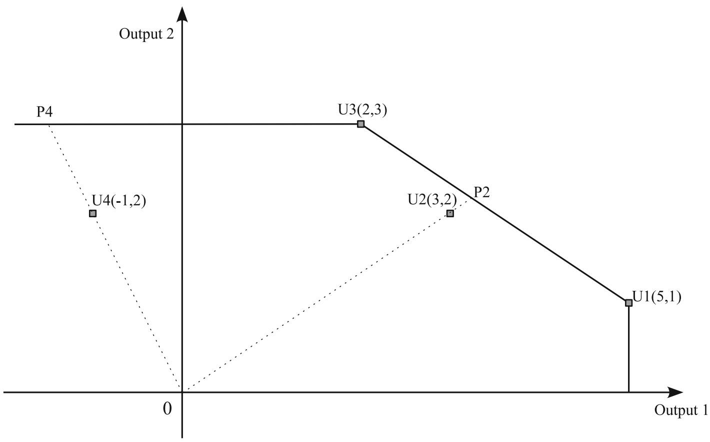


Fig. 8.5 Efficient frontier of the example with DMUs U_1, U_2, U_3, U_4 in the case $U_4 = (-1, 2)$ (normalized output values)

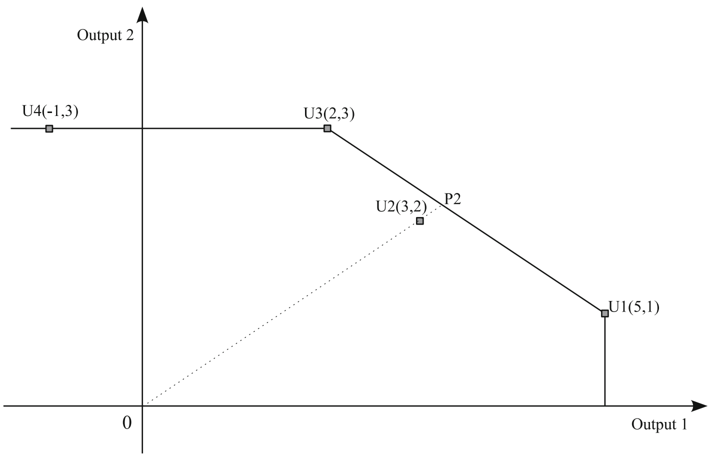


Fig. 8.6 Efficient frontier of the example with DMUs U_1, U_2, U_3, U_4 in the case $U_4 = (-1, 3)$ (normalized output values)

As can be seen, in the cases with $a \leq 3$ the inclusion of U_4 in the analysis does not modify the part of the efficient frontier which envelops U_1, U_2, U_3: this part is exactly the same as in the case without U_4; this is important because it entails that the efficiency scores of U_1, U_2, U_3 does not change either. For $a < 3$ (see Fig. 8.5) U_4 does not lie on the efficient frontier and therefore it is not efficient, while for $a = 3$ U_4 reaches the efficient frontier, as shown in Fig. 8.6, and therefore it becomes efficient.

In the cases with $3 < a \leq 5$, represented in Fig. 8.7, the displacement of U_4 upwards does modify the efficient frontier; however this shift does not alter the

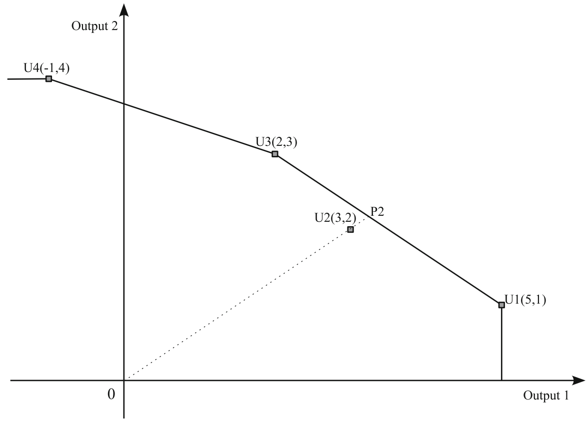


Fig. 8.7 Efficient frontier of the example with DMUs U_1, U_2, U_3, U_4 in the case $U_4 = (-1, 4)$ (normalized output values)

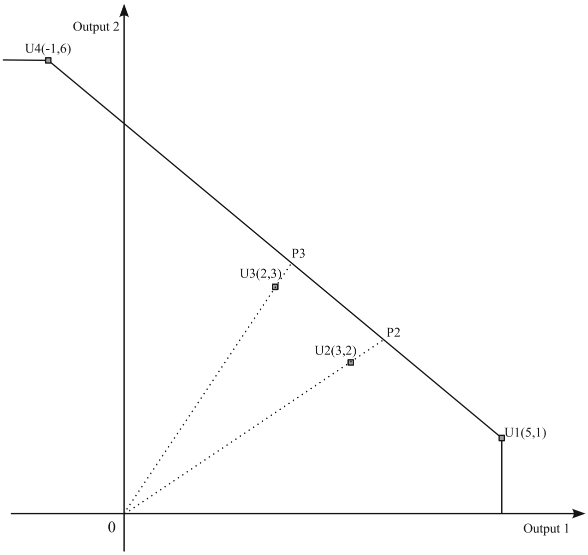


Fig. 8.8 Efficient frontier of the example with DMUs U_1, U_2, U_3, U_4 in the case $U_4 = (-1, 6)$ (normalized output values)

section of the efficient frontier that determines the efficiency scores of U_1, U_2, U_3, so that their performance measures do not change. On the other hand, for $a \geq 3$ U_4 lies on the efficient frontiers and hence it is efficient.

Figure 8.8 shows that for $a > 5$ the raising of U_4 moves the efficient frontier away from DMUs U_2 and U_3, causing a worsening of their efficiency scores; this shift makes U_3 become inefficient. Hence, a sufficiently high value of the second output can compensate for the negative value of the first output, in such a way as to make U_4 become efficient when the value of the second output is high enough.

On the other hand, let us keep the values of both the input and the second output constant while decreasing the value of the first (negative) output. In particular, let us analyze the behavior of the efficiency score of U_4 as the value of the negative

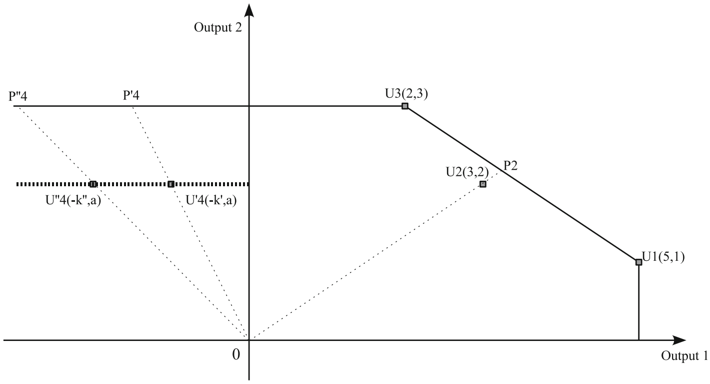


Fig. 8.9 Distance from the efficient frontier of DMU $U_4(-k,a)$ for different values of k and $a=2$ (normalized output values)

output worsens. In such a case, a good performance measure should exhibit a decreasing efficiency score for U_4 as the negative output value worsens. However, this is not what happens.

Actually, let us consider $U_4(-k,a)$ for $0<a\leq 3$ as $k>0$ increases; Fig. 8.9 shows the situation for $k'=1$ and $k''=2$, and $a=2$. It is easy to see that the Cartesian coordinates of the virtual unit P_4 on the efficient frontier are the following

$$P4\left(-\frac{3k}{a},3\right), \tag{8.16}$$

so that the DEA efficiency score of U_4 turns out to be constant and equal to

$$\frac{dist\,(O,U_4)}{dist\,(O,P_4)}=\frac{\sqrt{(-k)^2+a^2}}{\sqrt{\left(-\frac{3k}{a}\right)^2+3^2}}=\frac{a}{3}, \tag{8.17}$$

no matter the value of the first output. This means that the efficiency measure of U_4 is the same for all values of the first (negative) output, and thus the value of the second output (besides that of the input) is the only thing that matters.

In the context of the measurement of the performance of mutual funds, this fact has an unrealistic consequence, that does not satisfy the usual economic assumptions on the investors preferences. Actually, if the first output represents the average rate of return of the mutual fund (or its average excess return), this entails that when the average rate of return is negative, its value is indifferent for investors, whether it is only slightly less than zero or it entails a heavy loss. In this case only the value of the second output would be relevant, which is clearly in contrast with the economic principle that, all other things equal, a higher expected value is always preferred.

On the other hand, if we consider a single output that may take negative values (and any number of inputs), things are not better. Actually, it can be easily seen from the CCR-O model

$$\min_{u, v_1, \ldots, v_p} \sum_{i=1}^{p} v_i x_{ij_0} \tag{8.18}$$

s.t.

$$uy_{j_0} = 1 \tag{8.19}$$

$$\sum_{i=1}^{p} v_i x_{ij} \geq uy_j \qquad j = 1, 2, \ldots, n \tag{8.20}$$

$$u, v_1, v_2, \ldots, v_p \geq \varepsilon, \tag{8.21}$$

that if y_{j_0} is negative a solution does not even exist; in such a case the feasible region is empty, since the first constraint (8.19) has no solution with $u \geq \varepsilon$.

Things only partially improve introducing variable returns to scale with a BCC model. Indeed, a BCC-I model is translation invariant with respect to outputs (see Pastor 1996), which means that the DEA efficiency measure is invariant for translations of the original output values consequent to an addition of a constant to the original data. Therefore, with a BCC-I model the problem of negative returns (or excess returns) can be easily overcome by adding a suitable constant (greater than the absolute value of the lowest mean return) to the negative output; for an overview on this subject see also Pastor and Ruiz (2007).

However, since the performance results depend on the orientation of the model, the orientation should be carefully chosen on the basis of financial considerations. In our opinion, the most appropriate orientation for the assessment of the performance of mutual funds is the output orientation, since investors usually seek to maximize the value of mean returns and (eventual) other output variables without increasing the value of the input variables. But the BCC-O model is not translation invariant with respect to outputs, so that with this model we face drawbacks analogous to those encountered with CCR models; for further remarks see for instance Silva Portela et al. (2004).

On the other hand, we could try to adopt a suitable DEA model which is translation invariant (Ali and Seiford 1990; Lovell and Pastor 1995). In particular, a well known DEA model with such a property is the additive model, and actually this model is often used in order to tackle the problem of negative data in DEA analyses. However, an additive DEA model discriminates between efficient and inefficient DMUs, but it cannot gauge the depth of eventual inefficiencies: indeed, the efficiency measure given by an additive model does not provide a radial efficiency measure such as that given by the basic CCR and BCC models (for the additive model see e.g. Cooper et al. 2011), and its financial interpretation is far from being straightforward.

Another approach, the range directional model proposed in Silva Portela et al. (2004), treats the problem of negative data in DEA models by modifying the efficiency measure used, but neither this approach is directly connected to radial efficiency. Other contributions recently appeared in the literature propose different approaches which modify, in one way or another, the efficiency measure. Among them, we find the semi-oriented radial measure used by Emrouznejad et al. (2010), a generalised proportional distance function used by Kerstens and Van de Woestyne (2011), a shortage function adopted in Kerstens et al. (2012), a probabilistic characterization of directional distances devised by Simar and Vanhems (2012); Cheng et al. (2013) also proposes a way to deal with negative data.

Notwithstanding, these methodologies make the efficiency measures difficult to interpret, especially from a financial point of view. If we aim to assert the validity of a DEA efficiency measure for the evaluation of the performance of mutual funds and wish to spread its adoption in the financial practice, these approaches may well be met with distrust in the financial world. For this reason, we need a model which is at the same time simple, financially meaningful and able to deal with both positive and negative returns.

Actually, only some of the DEA models used in the literature to assess the performance of mutual funds can cope with the presence of negative mean returns. Table 8.10 reports the papers on mutual and hedge funds that explicitly tackle this problem and highlights the different solutions that they adopt to allow for the presence of negative returns in the DEA models.

In addition, as can be seen looking at Tables 8.1, 8.2, 8.3, 8.4, 8.5, 8.6, and 8.7, even without making any remark on the problem of negative data, some of the papers use either an input oriented BCC model or an additive model which are able to handle the problem.

On the other hand, let us point out that the presence of negative mean returns poses a problem not only in DEA modeling. Also most of the traditional indicators used to evaluate the performance of mutual funds run into serious problems that prevent a sensible usage in the presence of negative mean returns, as will be discussed in Sect. 8.5.

8.4 A DEA Model for the Performance Assessment in Periods of Financial Crisis

We adopt a performance measure recently proposed in Basso and Funari (2014a) which is simple to implement, meaningful from a financial point of view and able to deal with the negative returns characterising the periods of financial crisis. This model is focused on the objectives of investors and takes the point of view of a representative investor who has to pick the best mutual fund in the set $\{1, 2, \ldots, n\}$ of mutual funds analyzed.

Table 8.10 Negative data

Paper	Methodology adopted to allow for negative returns
Basso A., Funari S. (2005b)	Uses the mean capitalization factor, which is always positive, instead of mean return
Basso A., Funari S. (2008)	Uses the mean capitalization factor, which is always positive, instead of mean return
Basso A., Funari S. (2014a)	Uses the final value of the investment, which is always positive, instead of mean return
Basso A., Funari S. (2014b)	Uses the final value of the investment, which is always positive, instead of mean return
Basso A., Funari S. (2014c)	Uses the final value of the investment, which is always positive, instead of mean return
Basso A., Funari S. (2014d)	Uses the mean capitalization factor, which is always positive, instead of mean return
Gregoriou G.N., Zhu J. (2005)	Uses a model that is translation invariant with respect to outputs (BCC-I)
Hu J.-L., Yu H.-E., Wang Y.-T. (2012)	Uses a model that is translation invariant with respect to outputs (BCC-I)
Kumar U.D., Roy A., Saranga H., Singal K. (2010)	Uses slack-based models
Lozano S., Gutiérrez E. (2008a)	Uses an additive model
Lozano S., Gutiérrez E. (2008b)	Uses an additive model
Simar L., Vanhems A., Wilson P.W. (2012)	Uses a model with directional distance and an input oriented VRS model
Tavakoli Baghdadabad M.R., Noori Houshyar A. (2014)	Uses the mean capitalization factor, which is always positive, instead of mean return

Let K_j be the initial payout required by fund j to start with an initial capital equal to 1, net of the initial fee c_{Ij}:

$$K_j = \frac{1}{1 - c_{Ij}} \qquad j = 1, 2, \ldots, n. \tag{8.22}$$

This is the first input variable of the model; in addition, we consider also one or more risk measures $\{q_{1j}, q_{2j}, \ldots, q_{hj}\}$ that may shed light on different features of the financial risk of fund j. For example, we may include widely used measures, such as the historical volatility σ_j (the standard deviation of the returns of fund j) and the β-coefficient β_j (the ratio of the covariance between fund j and the market returns to the variance of the market return), but also some of the other measures proposed in the recent literature (see the review presented in Sect. 8.2).

As for the outputs, we consider a single output model in which the output variable is the final value M_j of the investment, net of the exit fee c_{Ej}; more precisely, M_j is defined as follows:

$$M_j = (1 + R_j)^T (1 - c_{Ej}) \qquad j = 1, 2, \ldots, n, \tag{8.23}$$

where R_j denotes the mean rate of return of fund j in the holding period of length T considered, measured on an annual basis using the compound interest regime.

Equivalently, we may compute the final value M_j using the continuous law of interest; as a matter of fact, this is the choice adopted in the empirical analysis that is presented in Sect. 8.6. In such a case, R_j denotes the mean instantaneous rate of return, measured on an annual basis using the continuous compounding, and M_j can be written as:

$$M_j = \mathrm{e}^{R_j T}(1 - c_{Ej}) \qquad j = 1, 2, \ldots, n. \tag{8.24}$$

Note that the final values M_j computed with formulas (8.23) and (8.24) coincide, since the instantaneous rate of return used in formula (8.24) is equal to the natural logarithm of 1 plus the compound rate of return of formula (8.23). Note also that either of the two methods to compute the rate of return, compound or continuous capitalization, may be used in the computation of the risk measures $\{q_{1j}, q_{2j}, \ldots, q_{hj}\}$.

We point out that the output variable is a measure of the overall profitability of the investment and, as such, depends heavily on the choice of the holding period.

On the other hand, we have $M_j \geq 0 \, \forall \, j$, independently of the phase of the business cycle, so that the model can easily be used also in the presence of negative mean returns.

Let us also observe that this choice of the input and output variables enables the model to take into account the initial and exit fees without the need to include them directly as input variables. We may thus avoid the problem of assessing a fund as efficient only because it has low fees.

As we have seen in Sect. 8.2, in the literature on mutual funds we find instances of both models with constant returns to scale (CRS) and others with variable returns to scale (VRS). We formulate both a constant returns to scale model, which will be denoted by DEA-C, and a variable returns to scale model, denoted by DEA-V.

In order to write the DEA-C model, which is a "plain vanilla" CCR model, let us begin with its formulation as a fractional programming problem:

$$\max_{\{u, v_i\}} \quad \frac{uM_o}{v_1 K_o + \sum_{i=2}^{h+1} v_i q_{io}} \tag{8.25}$$

subject to

$$\frac{uM_j}{v_1 K_j + \sum_{i=2}^{h+1} v_i q_{ij}} \leq 1 \quad j = 1, 2, \ldots, n \tag{8.26}$$

$$u \geq \varepsilon, \tag{8.27}$$

$$v_i \geq \varepsilon \quad i = 1, 2, \ldots, h+1 \tag{8.28}$$

where u is the weight assigned to the final value M_j, v_1 is the weight assigned to the initial payout $K_j, v_2, v_3, \ldots, v_{h+1}$ are the weights assigned to the h risk measures and

ε is a non-Archimedean constant. As usual in the data envelopment analysis, the optimal value of the objective function (8.25) gives the efficiency score of mutual fund $o \in \{1, 2, \ldots, n\}$.

As is well known, model (8.25)–(8.28) can be transformed into an equivalent linear programming problem. We may distinguish the case $M_o = 0$ from the case $M_o > 0$. When $M_o = 0$ the efficiency score is clearly equal to 0. When $M_o > 0$ we can adopt the output orientation and write the following equivalent linear program:

$$\min_{\{u, v_i\}} \quad v_1 K_o + \sum_{i=2}^{h+1} v_i q_{io} \tag{8.29}$$

subject to

$$uM_o = 1 \tag{8.30}$$

$$-uM_j + v_1 K_j + \sum_{i=2}^{h+1} v_i q_{ij} \geq 0 \quad j = 1, 2, \ldots, n \tag{8.31}$$

$$u \geq \varepsilon \tag{8.32}$$

$$v_i \geq \varepsilon \quad i = 1, 2, \ldots, h+1 \tag{8.33}$$

In this single output model we may observe that constraint (8.30) makes constraint (8.32) redundant.

We may also consider the dual of problem (8.29)–(8.33):

$$\max \quad z_0 + \varepsilon s^+ + \sum_{i=1}^{h+1} \varepsilon s_i^- \tag{8.34}$$

subject to

$$M_o z_0 - \sum_{j=1}^{n} M_j \lambda_j + s^+ = 0 \tag{8.35}$$

$$\sum_{j=1}^{n} K_j \lambda_j + s_1^- = K_o \tag{8.36}$$

$$\sum_{j=1}^{n} q_{ij} \lambda_j + s_i^- = q_{io} \qquad i = 2, 3, \ldots, h+1 \tag{8.37}$$

$$\lambda_j \geq 0 \quad j = 1, 2, \ldots, n \tag{8.38}$$

$$s^+ \geq 0 \tag{8.39}$$

$$s_i^- \geq 0 \quad i = 1, 2, \ldots, h+1, \tag{8.40}$$

where z_0 is the dual variable associated with the equality constraint (8.30), λ_j are the dual variables associated with the mutual funds constraints (8.31) and s_1^+ and s_i^- are the dual variables connected with the output and input weight constraints (8.32) and (8.33), respectively.

It is known that one of the advantages of the DEA methodology is that it gives further information to the inefficient units, with the indication of the so called "virtual unit", which is a combination of efficient units (the "peers") that is efficient with the inefficient unit's weights. The financial interpretation of the virtual unit is interesting, since it may be seen as an efficient benchmark portfolio with a similar profile, which the inefficient fund can strive to imitate (see Basso and Funari 2001).

This benchmark portfolio is defined as the linear combination of the peers with coefficients given by the optimal values λ_j^* of the dual variables λ_j and has inputs

$$\sum_{j=1}^{n} \lambda_j^* K_j \tag{8.41}$$

and

$$\sum_{j=1}^{n} \lambda_j^* q_{ij} \qquad i = 2, 3, \ldots, h+1 \tag{8.42}$$

and output

$$\sum_{j=1}^{n} \lambda_j^* M_j. \tag{8.43}$$

Let us now consider a convexity constraint on the "production possibility set" (see Banker et al. 1984); in our analysis of mutual funds, this means that we restrict the set of benchmark portfolios that can be considered to the convex hull of the funds analyzed, so that only convex combinations of the existing funds are allowed. This is obtained by adding the following convexity constraint

$$\sum_{j=1}^{n} \lambda_j = 1 \tag{8.44}$$

to the dual program (8.34)–(8.40). We obtain in this way the corresponding BCC model with variable returns to scale, namely model DEA-V:

$$\max \quad z_0 + \varepsilon s^+ + \sum_{i=1}^{h+1} \varepsilon s_i^- \tag{8.45}$$

subject to

$$M_o z_0 - \sum_{j=1}^{n} M_j \lambda_j + s^+ = 0 \tag{8.46}$$

$$\sum_{j=1}^{n} K_j \lambda_j + s_1^- = K_o \tag{8.47}$$

$$\sum_{j=1}^{n} q_{ij} \lambda_j + s_i^- = q_{io} \quad i = 2, 3, \ldots, h+1 \tag{8.48}$$

$$\sum_{j=1}^{n} \lambda_j = 1 \tag{8.49}$$

$$\lambda_j \geq 0 \quad j = 1, 2, \ldots, n \tag{8.50}$$

$$s^+ \geq 0 \tag{8.51}$$

$$s_i^- \geq 0 \quad i = 1, 2, \ldots, h+1, \tag{8.52}$$

With the addition of constraint (8.44), the primal program associated to the dual program (8.45)–(8.52) requires an additional variable, free in sign, whose sign allows one to perform a local analysis of the returns to scale for the points on the efficient frontier, telling whether they are increasing, constant or decreasing (see Banker et al. 1984).

8.5 More Traditional Indicators of Mutual Fund Performance

Various traditional indicators often used in finance to assess the performance of mutual funds either refer to ratios of the expected value of the excess return over a measure of the risk of the investment or are derived from well known financial models such as the Capital Asset Pricing Model (CAPM). Here the excess return is defined as the difference between the return of the fund and the return of a riskless asset.

As a matter of fact, since the expected values are not known, it is usual to compare the performance of mutual funds over a past period of time by replacing the (ex ante) expected return and the (unknown) value of the risk measure with the mean return R_j and the value of the risk measure computed ex post on the historical data in the period considered.

The most popular indicator defined as a ratio is probably the Sharpe index (see Sharpe 1994). Let $r_{j1}, r_{j2}, \ldots, r_{jT}$ be the rates of return obtained by fund j in the periods $1, 2, \ldots, T$; for example, we could consider the monthly rates of return of the last 3 years. Analogously, let $r_{f1}, r_{f2}, \ldots, r_{fT}$ be the risk-free rates of return in the same periods. The Sharpe index can be computed as follows:

$$I_{j,Sharpe} = \frac{R_j - R_f}{\sqrt{\mathrm{Var}[r_j - r_f]}}, \tag{8.53}$$

where R_j and R_f are the mean return of fund j and the mean risk-free rate in the holding period $[0, T]$, respectively, and $\mathrm{Var}[r_j - r_f]$ is the variance of the differences $r_{jt} - r_{ft}$.

Note that the Sharpe index, as well as the other traditional indicators, does not take into account the initial and exit fees required by the mutual fund.

The standard deviation used as risk measure in the Sharpe index is appropriate when the fund returns are normally distributed and the investor does not possess other risky assets.

When the investor possesses a well diversified portfolio of assets, a more suitable risk measure is the beta coefficient β_j; this is the risk measure used by the Treynor index (see Treynor 1965), which is defined as the ratio:

$$I_{j,Treynor} = \frac{R_j - R_f}{\beta_j}. \tag{8.54}$$

When the fund returns are not normally distributed, risk measures other than the standard deviation may better describe the risk. Among them, it is worth citing the downside risk DR, defined as the lower semi-deviation of the returns from a target value m (also called minimum acceptable return, Sortino and van der Meer 1991):

$$\mathrm{D}R_j = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\min[r_{jt} - m, 0])^2}. \tag{8.55}$$

This measure of risk translates the statement of fact that investors consider as unfavourable only the returns lower than the target value, not those that exceed it.

The downside risk is used by the Sortino index to define another risk adjusted indicator which is similar to the Sharpe ratio but measures the risk with the downside risk:

$$I_{j,Sortino} = \frac{R_j - m}{\mathrm{D}R_j}. \tag{8.56}$$

When the target is not chosen as a fix predetermined value but as the return of a benchmark, such as for example the return of a riskless asset, the Sortino index may be adjusted by computing the downside risk as follows:

$$\mathrm{D}R_j = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\min[r_{jt} - r_{ft}, 0])^2} \tag{8.57}$$

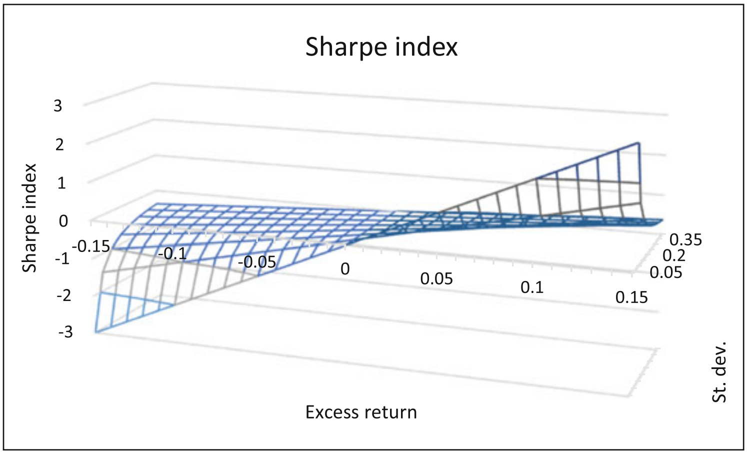


Fig. 8.10 Behavior of the Sharpe ratio as the excess return and the standard deviation vary

and writing:

$$I_{j,Sortino} = \frac{R_j - R_f}{\mathrm{D}R_j}. \tag{8.58}$$

On the other hand, for the funds that exhibit a negative mean excess return, these traditional indicators can give misleading results.

The situation is well depicted in Fig. 8.10 for the Sharpe ratio, but the same drawback is exhibited also by the Treynor and Sortino indices. This figure shows the behavior of the Sharpe ratio as the excess return $R_j - R_f$ and the standard deviation $\sqrt{\mathrm{Var}[r_j - r_f]}$ vary.

It can be noticed that when the excess return is positive, the value of the Sharpe index decreases with the standard deviation, as we expect for a performance indicator. However, when the excess return is negative, the Sharpe index exhibits the opposite behavior as its value increases with the standard deviation. This means that if we compare two mutual funds with a same negative value of the excess return (hence with a rate of return lower than the riskless interest rate), the Sharpe index leads to choose the fund with the highest standard deviation, i.e. that with the highest risk.

Other traditional performance indicators derive from the theory of the well known Capital Asset Pricing Model. The progenitor of these indicators, and still the most widely used among them, is Jensen's alpha index (see Jensen 1968), which measures the portfolio performance through the intercept α_j of the CAPM regression line and can be computed as follows:

$$\alpha_j = (R_j - R_f) - \beta_j(R_m - R_f), \tag{8.59}$$

where R_m is the mean return of the so called "market portfolio" (see Jensen 1968).

Note that the alpha index can be computed and effectively used regardless of the phase of the business cycle. A positive value of α_j indicates that fund j outperformed the market portfolio in the holding period considered, while a negative value denotes that the management of fund j was not able to obtain returns on the line to the market portfolio, once the risk is properly taken into account through the beta coefficient.

8.6 An Empirical Investigation on Different Holding Periods

In order to see how the DEA-V model handles the different phases of the business cycle and to compare the performance scores obtained with this model to the values obtained with the more traditional performance indices, we have carried out an empirical analysis on a large set of European mutual funds from different countries.

To this aim, the investigation is conducted on different holding periods, and the performance measures are computed first on a holding period of 7 years, ranging from 30/11/2006 to 30/11/2013. Then the analysis is performed on two different 3-year subperiods, the first and the last 3 years of the time horizon considered: the first holding period, $[0, 3]$, ranges from 30/11/2006 to 30/11/2009 and is characterized by a negative trend of the economy, the second, $[4, 7]$, ranges from 30/11/2010 to 30/11/2013 and is characterized by a positive trend.

The 312 mutual funds analyzed have been randomly chosen among the mutual funds domiciled in Western Europe; their distribution by country is presented in Table 8.11. Table 8.11 shows also the average values by country of the initial and exit fees and the mean instantaneous return measured on an annual basis.

The rate of returns $r_{j1}, r_{j2}, \ldots, r_{jT}$ have been computed as monthly total returns, accumulating the dividends paid in the holding period considered, and they have been calculated from the fund's net asset value (NAV) extracted from the Bloomberg database, expressed as a per-share amount in euros.

In the analysis we consider three different risk measures, namely the historical volatility σ_j, the β-coefficient β_j computed with respect to the STOXX Europe Total Market Index (TMI, which represents a market portfolio for the Western Europe region), and the downside risk DR_j computed with respect to the 12 month Euribor rate.

Table 8.11 shows all the average values by country and by holding period. Let us remark that the different length of the holding periods affects the final value M_j in different ways.

First, of course the mean return R_j is related to the holding period considered; with regard to this, note that the average value (computed on all the European funds considered) is heavily negative in the holding period $[0, 3]$ (-0.0987, i.e. around $-10\,\%$ per year), while it is strongly positive in the holding period $[4, 7]$ ($9\,\%$ per year); on the whole time interval $[0, 7]$ the overall return is low but positive

Table 8.11 Average values of the fundamental data for the European mutual funds analyzed by country of domicile and by holding period

Country	No. funds	c_I	c_E	β-coeff.	Standard dev.	Downside risk	Mean return	% negative mean ret.
Holding period [0,7]								
Austria	16	0.0400	0.0000	0.9499	0.1872	0.1532	−0.0070	62.5
Belgium	10	0.0300	0.0000	1.0239	0.1913	0.1533	0.0034	40.0
Germany	11	0.0223	0.0023	0.9028	0.1782	0.1405	0.0153	36.4
France	74	0.0307	0.0033	1.0195	0.1858	0.1466	0.0047	37.8
Great Britain	44	0.0232	0.0018	0.8787	0.1871	0.1441	0.0266	13.6
Ireland	11	0.0348	0.0109	0.9464	0.1831	0.1409	0.0312	9.1
Luxembourg	91	0.0231	0.0035	0.9205	0.1828	0.1431	0.0214	19.8
Norway	16	0.0064	0.0013	0.9992	0.1904	0.1487	0.0317	6.3
Sweden	17	0.0065	0.0012	0.9397	0.1826	0.1350	0.0361	11.8
Others	22	0.0111	0.0058	0.7765	0.1593	0.1230	0.0175	27.3
Western Europe	312	0.0238	0.0031	0.9381	0.1832	0.1432	0.0173	25.6
Holding period [0,3]								
Austria	16	0.0400	0.0000	1.0180	0.2356	0.2060	−0.1148	100.0
Belgium	10	0.0300	0.0000	1.0800	0.2383	0.2040	−0.1080	100.0
Germany	11	0.0223	0.0023	0.9388	0.2155	0.1864	−0.1183	100.0
France	74	0.0307	0.0033	1.0159	0.2203	0.1849	−0.0927	100.0
Great Britain	44	0.0232	0.0018	0.9397	0.2327	0.1972	−0.1329	97.7
Ireland	11	0.0348	0.0109	1.0187	0.2278	0.1901	−0.0979	90.9
Luxembourg	91	0.0231	0.0035	0.9703	0.2245	0.1899	−0.0893	96.7
Norway	16	0.0064	0.0013	1.0737	0.2425	0.2031	−0.0790	93.8
Sweden	17	0.0065	0.0012	0.9989	0.2243	0.1808	−0.0916	100.0
Others	22	0.0111	0.0058	0.7888	0.1893	0.1587	−0.0832	100.0
Western Europe	312	0.0238	0.0031	0.9774	0.2239	0.1889	−0.0987	98.1
Holding period [4,7]								
Austria	16	0.0400	0.0000	0.8479	0.1397	0.1014	0.0577	12.5
Belgium	10	0.0300	0.0000	0.9265	0.1421	0.1029	0.0743	10.0
Germany	11	0.0223	0.0023	0.8153	0.1369	0.0934	0.1051	0.0
France	74	0.0307	0.0033	1.0484	0.1537	0.1119	0.0793	1.4
Great Britain	44	0.0232	0.0018	0.7531	0.1345	0.0894	0.1261	0.0
Ireland	11	0.0348	0.0109	0.8168	0.1334	0.0888	0.1083	9.1
Luxembourg	91	0.0231	0.0035	0.8407	0.1381	0.0957	0.0866	8.8
Norway	16	0.0064	0.0013	0.8888	0.1347	0.0913	0.0866	6.3
Sweden	17	0.0065	0.0012	0.8604	0.1387	0.0912	0.0970	0.0
Others	22	0.0111	0.0058	0.7718	0.1276	0.0889	0.0784	9.1
Western Europe	312	0.0238	0.0031	0.8777	0.1404	0.0979	0.0900	5.1

Table 8.12 Correlation coefficients of the performance scores obtained with the DEA-V model using different risk measures (models DEA-V$_\beta$, DEA-V$_{\beta,\sigma}$ and DEA-V$_{\beta,DR}$) in the holding period $[0, 7]$

	DEA-V$_\beta$	DEA-V$_{\beta,\sigma}$	DEA-V$_{\beta,DR}$
DEA-V$_\beta$	1		
DEA-V$_{\beta,\sigma}$	0.991	1	
DEA-V$_{\beta,DR}$	0.996	0.997	1

(0.0173). From the last column of Table 8.11 we may observe the presence of a non negligible percentage of funds with a negative mean return even in the "good" periods; for example, in the holding period $[0, 7]$ this is observed in as many as one fund in four ($25.6\,\%$ of the mutual funds considered).

Secondly, the effect of the initial and exit fees on the yearly returns is lessened as the length of the holding period increases. On the other hand, the compounding of interests accentuates the effect of the yearly rate of return on the final value M_j as the length of the holding period increases.

The performance analysis is carried out computing the performance scores with the DEA-V model discussed in Sect. 8.4 and the traditional performance indicators presented in Sect. 8.5. As for the choice of the risk measures to include in the DEA-V model, we have considered primarily the β-coefficient, in order to take into account investors with a well diversified portfolio. In addition, we have also included a dispersion measure, in order to take into consideration also investors without a diversified portfolio.

Table 8.12 shows the values of the correlation coefficients of the performance scores obtained with the DEA-V model by considering only the β-coefficient (model DEA-V$_\beta$), β_j and the volatility σ_j (model DEA-V$_{\beta,\sigma}$) and β_j and the downside risk DR_j (model DEA-V$_{\beta,DR}$). All the values are close to 1 (greater than 0.99), so that their results are similar. In the following we analyze more in detail the results obtained with the DEA-V$_{\beta,DR}$ model.

The average results obtained with the different performance measures by country and by holding period are shown in Table 8.13, while Table 8.14 displays the values of their correlation coefficients.

We have seen in Sect. 8.5 that the usage of the traditional Sharpe, Treynor and Sortino indicators should be averted in the presence of negative mean excess returns since they can lead to misleading results. Well, as can be seen from Table 8.13, in the holding period $[0, 3]$ the mean excess return is negative for almost all mutual funds ($99.4\,\%$ of funds): this is evidently a period of financial crisis. As for the holding period $[4, 7]$, this is clearly a period of financial recovery; nonetheless, the mean excess return is still negative for a small number of funds ($6.4\,\%$). On the whole, in the 7-year holding period $[0, 7]$ a good $60\,\%$ (more precisely, $60.3\,\%$) of mutual funds exhibit a negative mean excess return, with marked differences among the countries.

From Table 8.14 we may see that the values of the correlation coefficients among the Sharpe, Treynor and Sortino ratios are very high (greater than 0.98 in

Table 8.13 Average values of the performance measures obtained with the DEA-$V_{\beta,DR}$ model and the Sharpe, Treynor, Sortino and alpha indices by country of domicile and by holding period; the mean excess return is also reported

Country	Mean excess ret.	% negative excess ret.	Sharpe index	Treynor index	Sortino index	Alpha index	DEA-$V_{\beta,DR}$ score
Holding period [0,7]							
Austria	−0.0299	93.8	−0.1482	−0.0291	−0.1811	0.0016	0.4689
Belgium	−0.0194	70.0	−0.0933	−0.0181	−0.1102	0.0146	0.5227
Germany	−0.0076	72.7	−0.0095	0.0019	−0.0051	0.0224	0.5995
France	−0.0181	79.7	−0.0903	−0.0163	−0.1098	0.0157	0.5243
Great Britain	0.0038	52.3	0.0229	0.0029	0.0368	0.0329	0.6346
Ireland	0.0083	18.2	0.0484	0.0098	0.0672	0.0397	0.6126
Luxembourg	−0.0014	54.9	−0.0012	0.0006	0.0073	0.0291	0.6158
Norway	0.0088	31.3	0.0521	0.0095	0.0716	0.0420	0.6736
Sweden	0.0132	29.4	0.0613	0.0103	0.0886	0.0444	0.7018
Others	−0.0053	63.6	−0.0227	0.0003	−0.0154	0.0204	0.6388
Western Europe	−0.0055	60.3	−0.0233	−0.0038	−0.0219	0.0256	0.5948
Holding period [0,3]							
Austria	−0.1512	100.0	−0.6553	−0.1515	−0.7534	0.0171	0.6362
Belgium	−0.1444	100.0	−0.6294	−0.1390	−0.7340	0.0341	0.6563
Germany	−0.1547	100.0	−0.6963	−0.1629	−0.8081	0.0005	0.6801
France	−0.1291	100.0	−0.5890	−0.1288	−0.7028	0.0389	0.6973
Great Britain	−0.1693	100.0	−0.7357	−0.1925	−0.8596	−0.0139	0.6475
Ireland	−0.1343	100.0	−0.5882	−0.1307	−0.6997	0.0341	0.6866
Luxembourg	−0.1257	97.8	−0.5853	−0.1354	−0.6902	0.0347	0.7282
Norway	−0.1154	100.0	−0.5138	−0.1139	−0.6139	0.0621	0.7725
Sweden	−0.1280	100.0	−0.6056	−0.1396	−0.7372	0.0372	0.7549
Others	−0.1196	100.0	−0.6559	−0.1677	−0.7904	0.0108	0.7938
Western Europe	−0.1351	99.4	−0.6188	−0.1450	−0.7319	0.0265	0.7077
Holding period [4,7]							
Austria	0.0453	12.5	0.3668	0.0601	0.5335	−0.0060	0.5849
Belgium	0.0618	10.0	0.4615	0.0707	0.6870	0.0058	0.6285
Germany	0.0927	0.0	0.7421	0.1493	1.1730	0.0434	0.7484
France	0.0669	4.1	0.4609	0.0695	0.6662	0.0035	0.6367
Great Britain	0.1137	0.0	0.8698	0.1717	1.3404	0.0682	0.7637
Ireland	0.0959	9.1	0.7404	0.1200	1.1884	0.0465	0.7042
Luxembourg	0.0742	11.0	0.5914	0.1020	0.9477	0.0234	0.6841
Norway	0.0742	6.3	0.6242	0.0936	0.9603	0.0205	0.6955
Sweden	0.0846	0.0	0.6585	0.1090	1.0361	0.0325	0.7178
Others	0.0660	9.1	0.5785	0.1121	0.9293	0.0193	0.6934
Western Europe	0.0776	6.4	0.5990	0.1040	0.9273	0.0245	0.6833

Table 8.14 Correlation coefficients of the performance measures obtained with the DEA-$V_{\beta,DR}$ model and the Sharpe, Treynor, Sortino and alpha indices by holding period

	Sharpe index	Treynor index	Sortino index	Alpha index	DEA-$V_{\beta,DR}$ score
Holding period [0,7]					
Sharpe index	1				
Treynor index	0.980	1			
Sortino index	0.996	0.983	1		
Alpha index	0.948	0.921	0.934	1	
DEA-$V_{\beta,DR}$ score	0.921	0.902	0.924	0.855	1
Holding period [0,3]					
Sharpe index	1				
Treynor index	0.856	1			
Sortino index	0.993	0.848	1		
Alpha index	0.928	0.933	0.926	1	
DEA-$V_{\beta,DR}$ score	0.625	0.481	0.584	0.522	1
Holding period [4,7]					
Sharpe index	1				
Treynor index	0.871	1			
Sortino index	0.977	0.904	1		
Alpha index	0.963	0.898	0.943	1	
DEA-$V_{\beta,DR}$ score	0.877	0.886	0.872	0.892	1

the holding period $[0, 7]$) and they are only slightly lower for the correlation between these ratios and the Jensen's alpha index (greater than 0.92 in the holding period $[0, 7]$). As for the scores obtained with the DEA-V model, it is interesting to note that the correlations involving the DEA-$V_{\beta,DR}$ score are higher with the Sharpe, Treynor and Sortino ratios (greater than 0.9 in $[0, 7]$) and slightly lower with the alpha index (0.855 in $[0, 7]$). This may be due to the fact that the DEA scores are expressed as ratios.

Figures 8.11 and 8.12 show in more detail the relationship between the DEA-$V_{\beta,DR}$ score and the value of the Sharpe, Treynor and Sortino indices; more precisely, these figures display the scatter plots of all the mutual funds analyzed with respect to these performance measures for the holding periods $[0, 3]$ (Fig. 8.11) and $[4, 7]$ (Fig. 8.12). The comparison between the scatter plots in the two holding periods highlights the higher dispersion of the performance results obtained for the mutual funds with the DEA model and a traditional indicator in the holding period of financial crisis.

A similar behavior is observed in Fig. 8.13 for the relationship between the DEA-$V_{\beta,DR}$ score and the value of the Jensen's alpha index in the two holding periods.

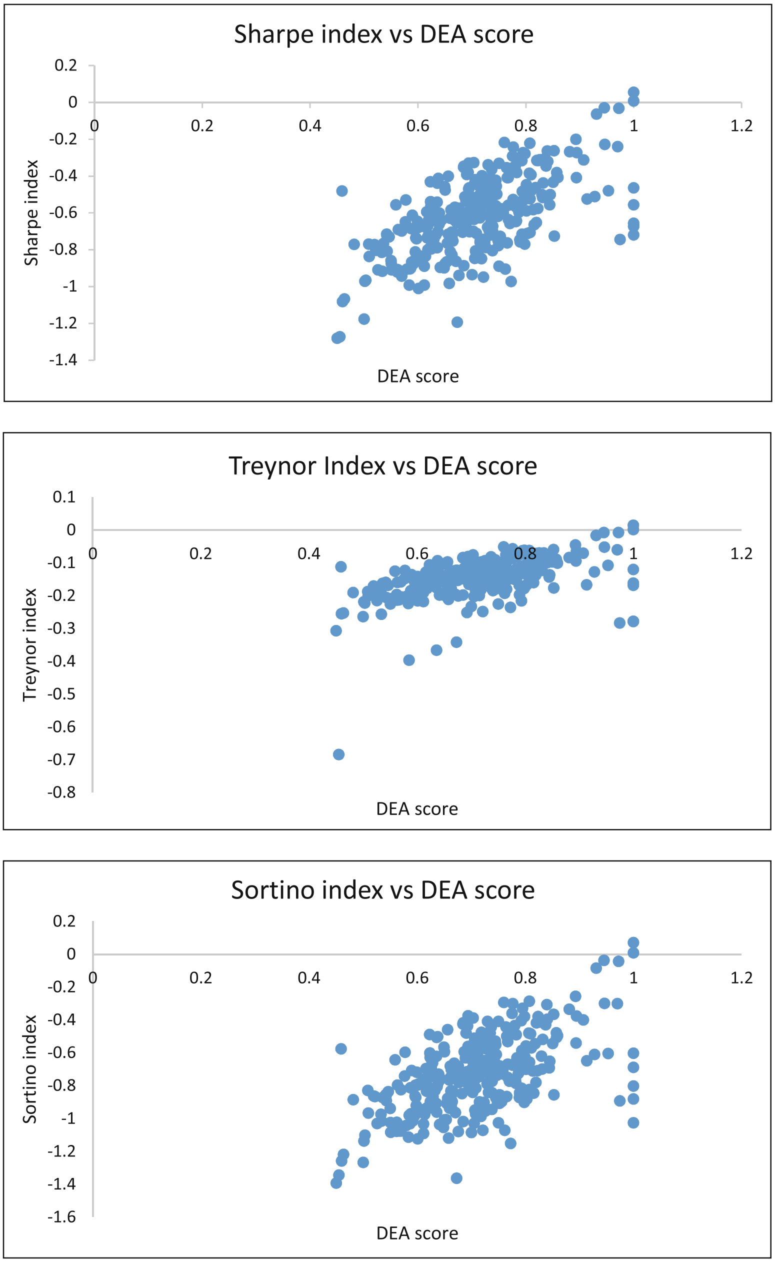


Fig. 8.11 Scatter plots of all mutual funds with respect to the DEA-$V_{\beta,DR}$ score and the Sharpe, Treynor and Sortino indices, respectively, in the holding period $[0, 3]$

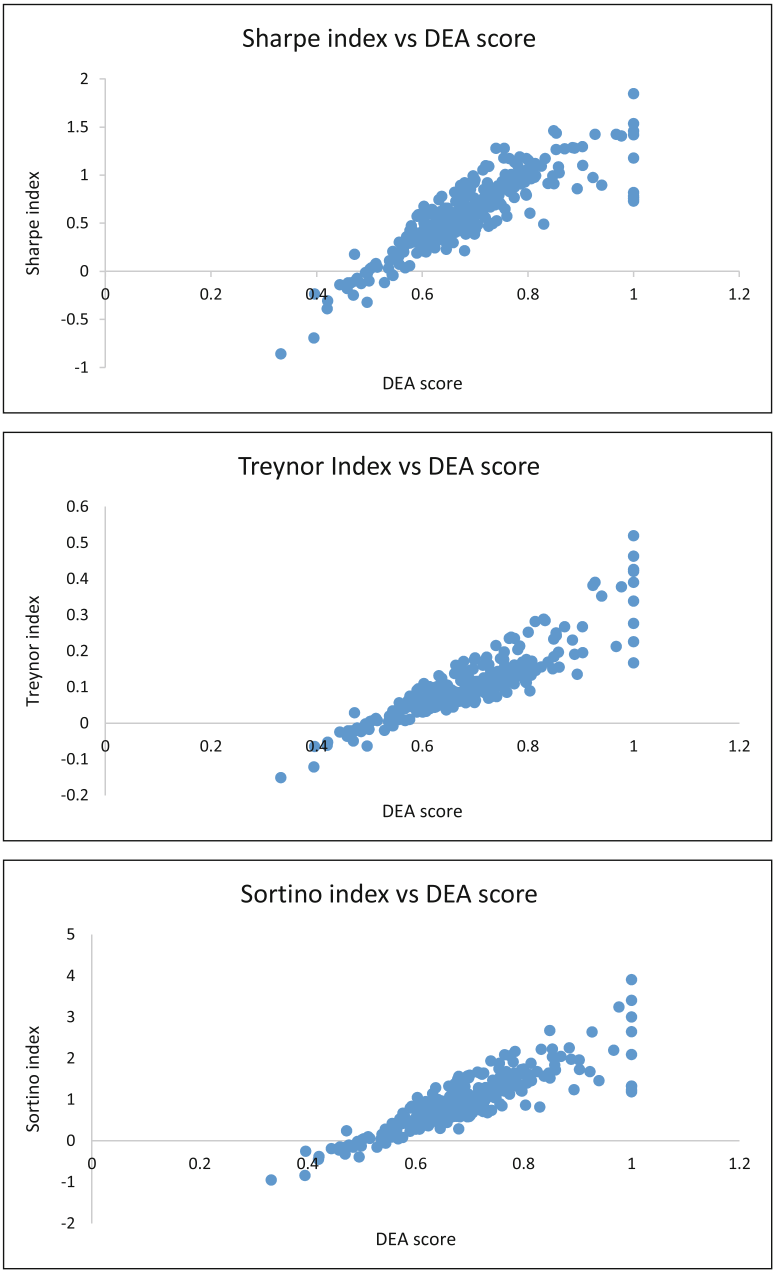
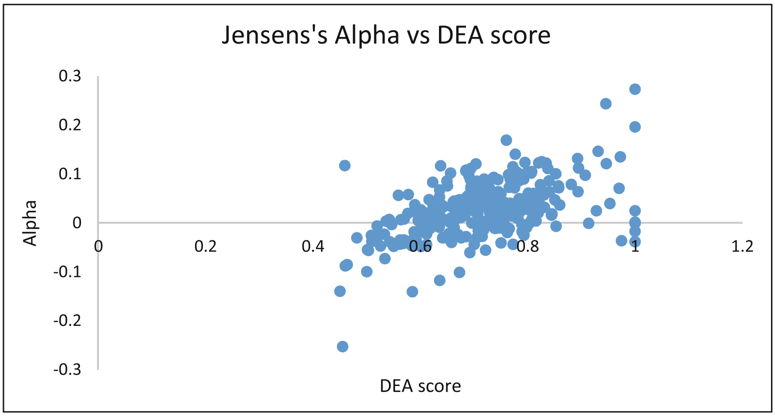


Fig. 8.12 Scatter plots of all mutual funds with respect to the DEA-V$_{\beta,DR}$ score and the Sharpe, Treynor and Sortino indices, respectively, in the holding period [4, 7]



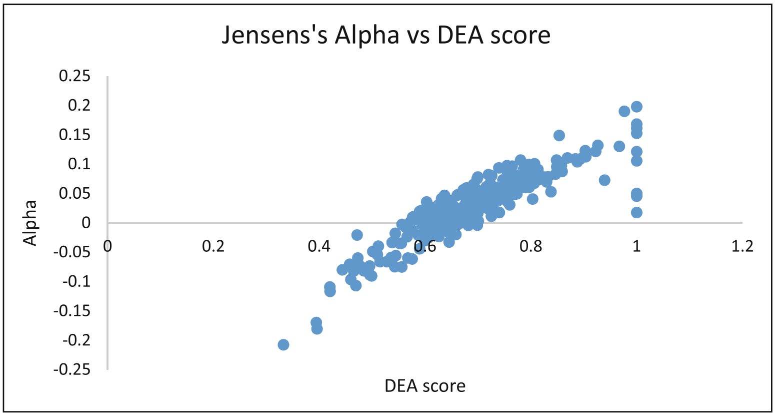


Fig. 8.13 Scatter plots of all mutual funds with respect to the DEA-$V_{\beta,DR}$ score and the Jensen's alpha index in the holding periods $[0, 3]$ and $[4, 7]$, respectively

References

Abdelsalam O, Duygun-Fethi M, Matallín JC, Tortosa-Ausina E (2014a) On the comparative performance of socially responsible and Islamic mutual funds. J Econ Behav Organ 103:S108–S128

Abdelsalam O, Duygun M, Matallín-Sáez JC, Tortosa-Ausina E (2014b) Do ethics imply persistence? The case of Islamic and socially responsible funds. J Bank Financ 40:182–194

Adamauskas S, Krusinskas R (2012) Behavioural finance efficiency under the influence of country's economic cycle. Inzinerine Ekon Eng Econ 23:327–337

Adeli O (2013) Measuring the efficiency of mutual funds. J Oper Res Appl (J Appl Math) 9:27–41

Alexakis P, Tsolas I (2011) Appraisal of mutual equity fund performance using data envelopment analysis. Multinatl Financ J 15:273–296

Ali AI, Seiford LM (1990) Translation invariance in data envelopment analysis. Oper Res Lett 9:403–405

Ali Asghar MJ-K, Afza T, Bodla MA (2013) Efficiency of the mutual funds in Pakistan. Middle-East J Sci Res 18:1055–1064

Ammann M, Moerth P (2008) Performance of funds of hedge funds. J Wealth Manag 2008:46–63
Anderson RI, Brockman CM, Giannikos C, McLeod RW (2004) A non-parametric examination of real estate mutual fund efficiency. Int J Bus Econ 3:225–238
Andreu L, Sarto JL, Vicente L (2014) Efficiency of the strategic style of pension funds: an application of the variants of the slacks-based measure in DEA. J Oper Res Soc 65:1886–1895
Babalos V, Caporale GM, Philippas N (2012) Efficiency evaluation of Greek equity funds. Res Int Bus Financ 26:317–333
Babalos V, Doumpos M, Philippas N, Zopounidis C (2015) Towards a holistic approach for mutual fund performance appraisal. Comput Econ 46:35–53
Baghdadabada MRT, Tanha FH, Halid N (2013) The efficiency evaluation of mutual fund managers based on DARA, CARA, IARA. J Bus Econ Manag 14:677–695
Bahrani R, Khedri N (2013) Evaluation of relative efficiency and performance of companies using data envelopment analysis (DEA) approach. Elixir Financ Manag 56:13299–13304
Banihashemi S, Sanei M (2013) Portfolio optimization and asset allocations using DEA and DEA/AHP. Ind J Sci Res 4:81–86
Banker R, Charnes A, Cooper WW (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis. Manag Sci 30:1078–1092
Barrientos A, Boussofiane A (2005) How efficient are pension fund managers in Chile? Rev Econ Contemp 9:289–311
Basso A, Funari S (2001) A data envelopment analysis approach to measure the mutual fund performance. Eur J Oper Res 135:477–492
Basso A, Funari S (2003) Measuring the performance of ethical mutual funds: a DEA approach. J Oper Res Soc 54:521–531
Basso A, Funari S (2005a) A generalized performance attribution technique for mutual funds. Cent Eur J Oper Res 13:65–84
Basso A, Funari S (2005b) Performance evaluation of ethical mutual funds in slump periods. Rendiconti per gli Studi Economici Quantitativi 2005:89–105
Basso A, Funari S (2008) DEA models for ethical and non ethical mutual funds. Math Methods Econ Financ 2:21–40
Basso A, Funari S (2014a) Constant and variable returns to scale DEA models for socially responsible investment funds. Eur J Oper Res 235:775–783
Basso A, Funari S (2014b) DEA models with a constant input for SRI mutual funds with an application to European and Swedish funds. Int Trans Oper Res 21:979–1000
Basso A, Funari S (2014c) The role of fund size and returns to scale in the performance of mutual funds. In: Perna C, Sibillo M (eds) Mathematical and statistical methods for actuarial sciences and finance. Springer, Cham, pp 21–25
Basso A, Funari S (2014d) Socially responsible mutual funds: an efficiency comparison among the European countries. In: Mathematical and statistical methods for actuarial sciences and finance. Springer, Cham, pp 69–79
Batra D, Batra G (2012) A DEA comparison of systematic and lump sum investment in mutual funds. Int J Comput Bus Res 3:1–10
Branda M (2011) Third-degree stochastic dominance and DEA efficiency-relations and numerical comparison. In: Dlouhý M, Skočdopolová V (eds) Proceedings of the 29th international conference on mathematical methods in economics. University of Economics in Prague, Jánská Dolina, pp 64–69
Branda M (2013a) Diversification-consistent data envelopment analysis with general deviation measure. Eur J Oper Res 226:626–635
Branda M (2013b) Diversification-consistent DEA-risk tests – solution techniques and an empirical comparison. In: Vojáčková H (ed) Proceedings of the 31th international conference on mathematical methods in economics. College of Polytechnics Jihlava, Jihlava, pp 77–82
Branda M (2013c) Reformulations of input-output oriented DEA tests with diversification. Oper Res Lett 41:516–520

Branda M, Kopa M (2012a) DEA-risk efficiency and stochastic dominance efficiency of stock indices. Czech J Econ Financ 62:106–124
Branda M, Kopa M (2012b) From stochastic dominance to DEA-risk models: portfolio efficiency analysis. In: Sakalauskas L, Tomasgard A, Wallace SW (eds) Proceedings of the international workshop on stochastic programming for implementation and advanced applications. Vilnius Gediminas Technical University, Vilnius, pp 13–18
Branda M, Kopa M (2014) On relations between DEA-risk models and stochastic dominance efficiency tests. Cent Eur J Oper Res 22:13–35
Brandouy O, Kerstens K, Van de Woestyne I (2015) Frontier-based vs.traditional mutual fund ratings: a first backtesting analysis. Eur J Oper Res 242:332–342
Briec W, Kerstens K, Lesourd JB (2004) Single-period Markowitz portfolio selection, performance gauging, and duality: a variation on the Luenberger shortage function. J Optim Theory Appl 120:1–27
Briec W, Kerstens K, Jokung O (2007) Mean-variance-skewness portfolio performance gauging: a general shortage function and dual approach. Manag Sci 53:135–149
Chang JF (2010) Applying DEA investment portfolio efficiency index and GA to the establishment of the fund of funds in Taiwan. Int J Organ Innov 2:225–249
Chang K-P (2004) Evaluating mutual fund performance: an application of minimum convex input requirement set approach. Comput Oper Res 31:929–940
Chen H-H (2008) Stock selection using data envelopment analysis. Ind Manag Data Syst 108:1255–1268
Chen Z, Lin R (2006) Mutual fund performance evaluation using data envelopment analysis with new risk measures. OR Spectr 28:375–398
Chen Y-C, Chiu Y-H, Li M-C (2011) Mutual fund performance evaluation - application of system BCC model. S Afr J Econ 79:1–16
Cheng G, Zervopoulos P, Qian Z (2013) A variant of radial measure capable of dealing with negative inputs and outputs in data envelopment analysis. Eur J Oper Res 225:100–105
Choi YK, Murthi BPS (2001) Relative performance evaluation of mutual funds: a non-parametric approach. J Bus Financ Account 28:853–876
Chu J, Chen F, Leung P (2010) ETF performance measurement-data envelopment analysis. In: International conference on service systems and service management (ICSSSM), 28–30 June 2010, pp 1–6
Cooper WW, Seiford LM, Zhu J (eds) Handbook on data envelopment analysis, 2nd edn. Springer, New York
Daraio C, Simar L (2006) A robust nonparametric approach to evaluate and explain the performance of mutual funds. Eur J Oper Res 175:516–542
Darling G, Mukherjee K, Wilkens K (2004) CTA performance evaluation with data envelopment analysis. In: Gregoriou GN, Karavas VN, Lhabitant F-S, Rouah F (eds) Commodity trading advisors: risk, performance analysis, and selection. Wiley, Hoboken, pp 79–104
de Guzman MPR, Cabanda EC (2005) Performance appraisal of Philippine mutual funds using DEA approach. Labuan Bull Int Bus Financ 3:65–77
Devaney M, Weber WL (2005) Efficiency, scale economies, and the risk/return performance of real estate investment trusts. J Real Estate Financ Econ 31:301–317
Dia M (2009) A portfolio selection methodology based on data envelopment analysis. Inf Syst Oper Res 47:71–79
Ding H, Zhou Z, Xiao H, Ma C, Liu W (2014) Performance evaluation of portfolios with margin requirements. Math Probl Eng 2014:1–8
Diz F, Gregoriou GN, Rouah F, Satchell SE (2004) Simple and cross-efficiency of CTAs using data envelopment analysis. In: Gregoriou GN, Karavas VN, Lhabitant F-S, Rouah F (eds) Commodity trading advisors: risk, performance analysis, and selection. Wiley, Hoboken, pp 129–147
Edirisinghe NCP, Zhang X (2007) Generalized DEA model of fundamental analysis and its application to portfolio optimization. J Bank Financ 31:3311–3335

Edirisinghe NCP, Zhang X (2010) Input/output selection in DEA under expert information, with application to financial markets. Eur J Oper Res 207:1669–1678
Einolf KW (2007) Building a socially responsible equity portfolio using data envelopment analysis. Philosophica 80:71–103
Eling M (2006) Performance measurement of hedge funds using data envelopment analysis. Financ Mark Portfolio Manag 20:442–471
Emrouznejad A, Anouze AL, Thanassoulis E (2010) A semi-oriented radial measure for measuring the efficiency of decision making units with negative data, using DEA. Eur J Oper Res 200:297–304
Galagedera DUA (2013) A new perspective of equity market performance. J Int Financ Mark Inst Money 26:333–357
Galagadera DUA, Silvapulle P (2002) Australian mutual fund performance appraisal using data envelopment analysis. Manag Financ 28:60–73
Glawischnig M, Sommersguter-Reichmann M (2010) Assessing the performance of alternative investments using non-parametric efficiency measurement approaches: is it convincing? J Bank Financ 34:295–303
Gnanasekar IF, Arul R (2013) Financial risk tolerance using data envelopment analysis-a case study. ZENITH Int J Bus Econ Manag Res 3:251–262
Gnanasekar IF, Arul R (2014) Data envelopment analysis: Jeopardy can never be defeat without taking financial risk tolerance. Indian J Appl Res 4:121–124
Gökgöz F (2010) Measuring the financial efficiencies and performances of Turkish funds. Acta Oecon 60:295–320
Gökgöz F, Çandarli D (2011) Data envelopment analysis: a comparative efficiency measurement for Turkish pension and mutual funds. Int J Econ Perspect 5:261–281
Gregoriou GN (2003) Performance appraisal of funds of hedge funds using data envelopment analysis. J Wealth Manag 5:88–95
Gregoriou GN (2006a) Optimisation of the largest US mutual funds using data envelopment analysis. J Asset Manag 6:445–455
Gregoriou GN (2006b) Trading efficiency of commodity trading advisors using data envelopment analysis. Derivatives Use Trading Regul 12:102–114
Gregoriou GN (2007) Efficiency of US mutual funds using data envelopment analysis. In: Gregoriou GN (ed) Performance of mutual funds: an international perspective. Palgrave Macmillan, Basingstoke, pp 152–167
Gregoriou GN, Chen Y (2006) Evaluation of commodity trading advisors using fixed and variable and benchmark models. Ann Oper Res 145:183–200
Gregoriou GN, Henry SC (2015) Undesirable outputs in commodities trading advisers: a data envelopment analysis approach. J Wealth Manag 17:85–92
Gregoriou GN, Zhu J (2005) Evaluating hedge fund and CTA performance: data envelopment analysis approach. Wiley, New York
Gregoriou GN, Zhu J (2007) Data envelopment analysis: a way to assess the efficiency of funds of hedge funds. J Portfolio Manag 33:120–132
Gregoriou GN, Rouah F, Sedzro K (eds) (2003) Performance evaluation of hedge funds. Beard Books, Washington
Gregoriou GN, Sedzro K, Zhu J (2005) Hedge fund performance appraisal using data envelopment analysis. Eur J Oper Res 164:555–571
Guillén GB (2008) Is efficiency the equivalent to a high rate of return for private pension funds? Evidence from Latin American countries. J CENTRUM Cathedra 1:109–117
Guo J, Ma C, Zhou Z (2012) Performance evaluation of investment funds with DEA and higher moments characteristics: financial engineering perspective. Syst Eng Proc 3:209–216
Hanafizadeh P, Khedmatgozar HR, Emrouznejad A, Derakhshan M (2014) Neural network DEA for measuring the efficiency of mutual funds. Int J Appl Decis Sci 7:255–269
Haslem JA, Scheraga CA (2003) Data envelopment analysis of Morningstar's large-cap mutual funds. J Invest 12:41–48

Haslem JA, Scheraga CA (2006) Data envelopment analysis of Morningstar's small-cap mutual funds. J Invest 15:87–92
Hsu C-M (2014) An integrated procedure for resolving portfolio optimization problems using data envelopment analysis, ant colony optimization and gene expression programming. Int J Comput Sci Bus Inform 9:45–65
Hsu C-S, Lin J-R (2007) Mutual fund performance and persistence in Taiwan: a non-parametric approach. Serv Ind J 27:509–523
Hu J-L, Chang T-P (2008) Decomposition of mutual fund underperformance. Appl Financ Econ Lett 4:363–367
Hu J-L, Yu H-E, Wang Y-T (2012) Manager attributes and fund performance: evidence from Taiwan. J Appl Financ Bank 2:85–101
Huang TH, Leu YH (2014) A mutual fund investment method using fruit fly optimization algorithm and neural network. Appl Mech Mater 571–572:318–325
Huang C-Y, Chiou C-C, Wu T-H, Yang S-C (2015) An integrated DEA-MODM methodology for portfolio optimization. Oper Res 15:115–136
Ismail MKA, Salamudin N, Rahman NMNA, Kamaruddin BH (2012) DEA portfolio selection in Malaysian stock market. In: 2012 international conference on innovation management and technology research (ICIMTR), 21–22 May 2012, pp 739–743
Jensen MC (1968) The performance of mutual funds in the period 1945–1964. J Financ 23:389–416
Jeong SO, Hoss A, Park C, Kang K-H, Ryu Y (2013) Portfolio selection for socially responsible investment via nonparametric frontier models. Commun Stat Appl Methods 20:115–127
Joro T, Na P (2006) Portfolio performance evaluation in a mean-variance-skewness framework. Eur J Oper Res 175:446–461
Kadoya S, Kuroko T, Namatame T (2008) Contrarian investment strategy with data envelopment analysis concept. Eur J Oper Res 189:120–131
Kerstens K, Van de Woestyne I (2011) Negative data in DEA: a simple proportional distance function approach. J Oper Res Soc 62:1413–1419
Kerstens K, Mounir A, Van de Woestyne I (2011) Non-parametric frontier estimates of mutual fund performance using C- and L-moments: some specification tests. J Bank Financ 35:1190–1201
Kerstens K, Mounir A, Van de Woestyne I (2012) Benchmarking mean-variance portfolios using a shortage function: the choice od direction vector affects rankings! J Oper Res Soc 63:1199–1212
Khedmatgozar HR, Kazemi A, Hanafizadeh P (2013) Mutual fund performance evaluation: a value efficiency analysis approach. Int J Electron Financ 7:263–280
Kumar UD, Roy A, Saranga H, Singal K (2010) Analysis of hedge fund strategies using slack-based DEA models. J Oper Res Soc 61:1746–1760
Kumar Singh A, Sahu R, Bharadwaj S (2010) Portfolio evaluation using OWA-heuristic algorithm and data envelopment analysis. J Risk Financ 11:75–88
Lamb JD, Tee K-H (2012a) Data envelopment analysis models of mutual funds. Eur J Oper Res 216:687–696
Lamb JD, Tee K-H (2012b) Resampling DEA estimates of investment fund performance. Eur J Oper Res 223:834–841
Lamb JD, Tee K-H (2012c) Data envelopment analysis (DEA) integrated risk assessment technique on hedge funds investment: theory and practical application. In: Risk assessment and management. Academy Publish, Wyoming, pp 73–84
Lim S, Oh KW, Zhu J (2014) Use of DEA cross-efficiency evaluation in portfolio selection: an application to Korean stock market. Eur J Oper Res 236:361–368
Lin R (2009) Mutual fund performance dynamic evaluation using data envelopment windows analysis. In: International conference on management and service science, pp 1–4
Lin R, Chen Z (2008) New DEA performance evaluation indices and their applications in the American fund market. Asia Pac J Oper Res 25:421

Lopes A, Lanzer E, Lima M, da Costa N Jr (2008) DEA investment strategy in the Brazilian stock market. Econ Bull 13:1–10
Lovell CAK, Pastor JT (1995) Units invariant and translation invariant DEA models. Oper Res Lett 18:147–151
Lozano S, Gutiérrez E (2008a) Data envelopment analysis of mutual funds based on second-order stochastic dominance. Eur J Oper Res 189:230–244
Lozano S, Gutiérrez E (2008b) TSD-consistent performance assessment of mutual funds. J Oper Res Soc 59:1352–1362
Majid MSA, Maulana H (2010) Assessing performance of mutual funds in Indonesia. J Econ Coop Dev 31:49–76
Majid MSA, Maulana H (2012) A comparative analysis of the productivity of Islamic and conventional mutual funds in Indonesia: data envelopment analysis (DEA) and general least square (GLS) approaches. Gadjah Mada Int J Bus 14:183
Malhotra DK, Malhotra R (2012) Using data envelopment analysis framework to evaluate mutual fund efficiency. In: 2012 Northeast decision sciences institute conference proceedings, pp 137–143
Margaritis DM, Otten R, Tourani-Rad A (2007) New Zealand equity fund performance appraisal: a non-parametric approach. In: Gregoriou G (ed) Performance of mutual funds: an international perspective. Palgrave-McMillan, New York, pp 17–30
Matallín-Sáez JC, Soler-Domínguez A, Tortosa-Ausina E (2014) On the informativeness of persistence for evaluating mutual fund performance using partial frontiers. Omega Int J Manag Sci 42:47–64
McMullen PR, Strong RA (1998) Selection of mutual funds using data envelopment analysis. J Bus Econ Stud 4:1–12
Medeiros Garcia MT (2010) Efficiency evaluation of the Portuguese pension funds management companies. J Int Financ Mark Inst Money 20:259–266
Morey MR, Morey RC (1999) Mutual fund performance appraisals: a multi-horizon perspective with endogenous benchmarking. Omega Int J Manag Sci 27:241–258
Murthi BPS, Choi YK, Desai P (1997) Efficiency of mutual funds and portfolio performance measurement: a non-parametric approach. Eur J Oper Res 98:408–418
Pai V (2012) Risk budgeted portfolio optimization using an extended ant colony optimization metaheuristic. Int J Appl Metaheuristic Comput 3:25–42
Pastor JT (1996) Translation invariance in data envelopment analysis: a generalization. Ann Oper Res 66:93–102
Pastor JT, Ruiz JL (2007) Variables with negative values in DEA. In: Zhu J, Cook WD (eds) Modeling data irregularities and structural complexities in data envelopment analysis. Springer, New York, pp 63–84
Pätäri E, Leivo T, Honkapuro S (2010) Enhancement of value portfolio performance using data envelopment analysis. Stud Econ Financ 27:223–246
Pätäri E, Leivo T, Honkapuro S (2012) Enhancement of equity portfolio performance using data envelopment analysis. Eur J Oper Res 220:786–797
Pendaraki K (2012) Mutual fund performance evaluation using data envelopment analysis with higher moments. J Appl Financ Bank 2:97–112
Pérez-Gladish B, Rodríguez MP, M'zali B, Lang P (2013) Mutual funds efficiency measurement under financial and social responsibility criteria. J Multi-Criteria Decis Anal 20:109–125
Pestana Barros C, Medeiros Garcia MT (2006) Performance evaluation of pension funds management companies with data envelopment analysis. Risk Manag Insur Rev 9:165–188
Powers J, McMullen PR (2000) Using data envelopment analysis to select efficient large market cap securities. J Bus Manag 7:31–42
Prasanna PK (2012) Performance of exchange-traded funds in India. Int J Bus Manag 7:122–143
Premachandra I, Powell JG, Shi J (1998) Measuring the relative efficiency of fund management strategies in New Zealand using a spreadsheet-based stochastic data envelopment analysis model. Omega Int J Manag Sci 26:319–331
Premachandra I, Zhu J, Watson J, Galagadera DUA (2012) Best-performing US mutual fund family from 1993 to 2008: evidence from a novel two-stage DEA model for efficiency decompostition. J Bank Financ 36:3302–3317

Rubio JF, Hassan MK, Merdad HJ (2012) Nonparametric performance measurement of international and Islamic mutual funds. Account Res J 25:208–226

Saad NM, Majid MSA, Kassim S, Hamid Z, Yusof RM (2010) A comparative analysis of the performance of conventional and Islamic unit trust companies in Malaysia. Int J Manag Financ 6:24–47

Sahoo BK, Meera E (2008) A comparative application of alternative DEA models in selecting efficient large cap market securities in India. Int J Manag Perspect 2:62–74

Sedzro K (2009) New evidence on hedge fund performance measures. Int Bus Econ Res J 8:95–105

Sengupta JK (2003) Efficiency tests for mutual fund portfolios. Appl Financ Econ 13:869–876

Sharpe WF (1994) The sharpe ratio. J Portfolio Manag 21:49–58

Silva Portela MCA, Thanassoulis E, Simpson G (2004) Negative data in DEA: a directional distance approach applied to bank branches. J Oper Res Soc 55:1111–1121

Simar L, Vanhems A (2012) Probabilistic characterization of directional distances and their robust versions. J Econom 166:342–354

Simar L, Vanhems A, Wilson PW (2012) Statistical inference for DEA estimators of directional distances. Eur J Oper Res 220:853–864

Soongswang A, Sanohdontree Y (2011a) Equity mutual fund: performances, persistence and fund rankings. J Knowl Manag Econ Inf Technol 6:1–27

Soongswang A, Sanohdontree Y (2011b) Open-ended equity mutual funds. Int J Bus Soc Sci 2:127–136

Sortino FA, van der Meer R (1991) Downside risk. J Portfolio Manag 17:27–31

Tarim SA, Karan MB (2001) Investment fund performance measurement using weight restricted data envelopment analysis: an application to the Turkish capital market. Russ East Eur Financ Trade 37:64–84

Tavakoli Baghdadabad MR, Noori Houshyar A (2014) Productivity and efficiency evaluation of US mutual funds. Czech J Econ Financ (Finance a úvěr) 64:120–143

Tokic D (2012) Managed futures for long-term investors: a DEA ranking analysis. Aust Econ Rev 45:422–440

Treynor JL (1965) How to rate management of investment funds. Harv Bus Rev 43:63–75

Tsolas IE (2011) Natural resources exchange traded funds: performance appraisal using DEA modeling. J CENTRUM Cathedra 4:250–259

Tsolas IE (2014) Precious metal mutual fund performance appraisal using DEA modeling. Resour Policy 39:54–60

Tsolas IE, Charles V (2015) Green exchange-traded fund performance appraisal using slacks-based DEA models. Oper Res 15:51–77

Wilkens K, Zhu J (2001) Portfolio evaluation and benchmark selection: a mathematical programming approach. J Altern Invest 4:9–19

Zamani L, Beegam R, Borzoian S (2014) Portfolio selection using data envelopment analysis (DEA): a case of select Indian investment companies. Int J Curr Res Acad Rev 2:50–55

Zhao X, Shi J (2010) Evaluation of mutual funds using multi-dimensional information. Front Comput Sci China 4:237–253

Zhao X, Wang S-Y (2007) Empirical study on Chinese mutual funds' performance. Syst Eng Theory Pract 27:1–11

Zhao X, Yue W (2008) Coned context DEA model with application to mutual funds evaluation. In: IEEE international conference on industrial engineering and engineering management, pp 805–809

Zhao X, Yue W (2012) A multi-subsystem fuzzy DEA model with its application in mutual funds management companies' competence evaluation. Proc Comput Sci 1:2469–2478

Zhao X, Wang S, Lai KK (2011) Mutual funds performance evaluation based on endogenous benchmarks. Expert Syst Appl 38:3663–3670

Chapter 9
Formulating Management Strategy for International Tourist Hotel Using DEA

Shiuh-Nan Hwang and Te-Yi Chang

Abstract In the face of a highly competitive environment, it has long been considered important for a hotel to formulate a business competition strategy, strengthen corporate operations and upgrade quality of service. This paper is to discuss the management strategies formulation of the international tourist hotel industry in Taiwan based on the efficiency evaluation. First, data envelopment analysis (DEA), developed by Charnes et al. (Eur J Oper Res 2(6):429–444, 1978), is used to measure the relative managerial efficiency of the international tourist hotels in Taiwan. Through the Kruskal-Wallis, Taiwan's international tourist hotels are divided into different strategic group. Secondly, by way of factor analyses of industry strategies, a clear picture of the industry strategies and features are identified. Consequently, the appropriate management strategies for different market segments are discussed. The research shows that the efficiency of international tourist hotels is different due to their managerial styles. International tourist hotels can be categorized into three groups: international cooperation; domestic franchise; and independent. "Target market", "product differentiation", "the degree of vertical integration", "economics of scale", "location" and "competitive edges" is concluded as the key factors to the formation of the managerial strategies. These six variables are used to discuss the managerial strategies of each market segment. The result of this paper will provide useful information for future business management needs of managers. Also, they may serve as valuable reference to the relevant authority of tourism.

Keywords International tourist hotel • Efficiency evaluation • Management strategic • Data envelopment analysis

S.-N. Hwang (✉)
Department of Business Administration, Ming Chuan University, 250, Chung Shan N. Rd., Sec. 5, Taipei 111, Taiwan, ROC
e-mail: snhwang@mail.mcu.edu.tw

T.-Y. Chang
Graduate Institute of Travel and Tourism Management, National Kaohsiung University of Hospitality and Tourism, No. 1, Songhe Rd., Xiaogang Dist., Kaohsiung 812, Taiwan, ROC

J. Zhu (ed.), *Data Envelopment Analysis*, International Series in Operations Research & Management Science 238, DOI 10.1007/978-1-4899-7684-0_9

9.1 Introduction

Strategy formulation is the set of processes involved in creating the strategies of the organization. The starting point in formulating strategy is usually SWOT analysis. One must first measure the comparative performance of the entire industry, before one may understand one's strengths and weaknesses. At the same time, one has to evaluate an organization's opportunities and threats. Therefore, a number of frameworks have been developed for identifying the major strategic alternatives that organizations should consider when choosing their strategies.

The strategy types in the study of Miller and Snow (1978) as well as the basic competitive strategy presented by Porter (1980, 1991) emphasize the relationship between strategy and changes in the external environment, when some scholars have presented core competencies (Ansoff 1990; Prahalad and Hamel 1990) and competitive advantage (Day 1994), focusing on the internal qualities of the organization as well as the research on the fit between resources and competencies. Base on the abovementioned research viewpoints, to understand the differences between the characteristics of various resources and the efficiency of resource input and utilization is the key to the organization improving business performance and building competitive advantage.

International tourist hotel operations not only require enormous capital and a huge labor force, but also modern equipment and sound management skills to improve service quality and enhance business performance. Conner (1991) and Okavarrueta (1996) went as far as to emphasize that organizations need to have specific resources and skills to enhance business performance. Roth et al. (1995) further clarified that for the service industry to produce good performance, an important factor of competitive advantage is strategic marketing and the integration of resources and service capabilities. There is a need to study the resources put in by the organization as well as the efficiency aspect of the output to enhance business performance and create competitive advantage for international tourist hotels.

In addition, with intense industry competition, coming up with a clear competitive strategy is the key to strengthening the corporate business form and raising the quality of service. Strategy group analysis provides a way to understand industry's competitive structure. After being first presented by Hunt in 1972, many scholars have focused on studies related to strategy groups. The study hopes to use strategy group analysis to categorize the international tourist hotel industry into groups according to different strategy styles and study the differences between them. Through group analysis, hotel managers will be able to understand the competitive position the entire industry is in as well as the strategy style of high-performing industry groups; this will serve as strong reference for managers in their choice of strategies.

Therefore, taking off from a performance evaluation standpoint, this study prioritizes the implications of multiple inputs and outputs of resource utilization efficiency. It uses the Data Envelopment Analysis (DEA) presented by Charnes et al. (1978) to evaluate the relative managerial efficiency of Taiwan's international tourist hotels. Finally, based on the results of managerial efficiency and operational features of the hotel industry, this paper use the concept of strategy groups to divide Taiwan's international tourist hotels into different groups to study the business strategies appropriate for each group, creating competitive advantage and achieving the goal of sustainable business.

9.2 Research Problem

9.2.1 Issues

With the economic development and technological progress seen in recent years, barriers to international travel have gradually been eliminated. Moreover, with a boom in international trade, international exchanges have largely increased, contributing to a rapid growth in global travel and tourism. Tourism has not only become one of the largest sources of income for many countries but also has an effective means to stimulate global economic development.

Since Taiwan entered the martial law in 1949, the development of international tourism industries in Taiwan has encountered significant limitations. With the lifting of this law in 1985, tourist activities have grown and the number of tourist arrivals increased from 1.3 million in 1985 to 5.56 million in 2010. Revenues from travel and foreign currencies increased from US$919 million to US$8719 million in 2010. With the lifting of martial law in Taiwan, hotels have mushroomed nationwide. International hotel numbers have increased from 44 in 1985 to 68 in 2010.

To improve the ability to receive international travelers, hoteliers have aggressively joined international chain organizations. By bringing in the management skills, professional concepts, as well as management techniques and talents of these international chains, they hope to increase the number of international clients and expand their market base. These hoteliers also see a huge potential in Taiwan's tourism market.

In recent years, the number of business travelers has increased and newly operated hotels, such as the Hyatt, Westin, Prince, Sheraton, Shangrila, and Novotel, Crowne Plaze, Okura, Kagaya, W Hotel, Le Meridien have entered into the market. Amid fierce competition, the older hotels like the Grand Hotel, Ambassador Hotel, Mandarina Crown Hotel were refurbished to retain and attract new customers. Some hotels of independent operations joined franchise-chains, to increase management competencies, local hotel brands need to integrate innovation. Groups like the FIH Regent Group and the L'Hotel de Chine Group, to name a few, have been aggressively consolidating their brands to respond to the market's needs.

In the face of intense industry competition, the crucial issue is how to come up with a marketing competition strategy to strengthen the corporate operational form and raise the quality of service. Taiwan's hotels adopt two ways to respond to competition. First, they segment the market by targeting international travelers and domestic tourists so as to increase the sources of customers. Second, some hotels joined franchise-chains, outsourced management and acquired membership in international hotel associations to introduce international management systems and promoting managerial capacities.

However, what kind of strategy is more effective in the highly competitive environment? When formulating any strategy, is it necessary to have a sound knowledge of relative managerial efficiency of a given hotel relative to the entire industry? What are the factors that affect managerial efficiency in the industry? Which hotels serve as positive examples? All of these help evaluate one's strengths and weaknesses in formulating strategies.

Therefore, this paper discusses two issues:

1. What is the relative managerial efficiency of international tourist hotels today? Will managerial efficiency differ in different market conditions, sources of customers, room size and management style?
2. What is the strategic management factors affecting international tourist hotels? How is the management performance of each group of hotels? What are their business strategies?

To answer the above questions, this paper used data envelopment analysis (DEA) to measure the managerial efficiency of international hotels in Taiwan. Based on patterns of operation, hotels are classified into three large groups: independent operations, domestic chains and international chain operations. International chain operations are further subdivided into franchise chain, management contract and membership.

And strategic decision-making factors for the hotel industry is divided into six value activities: target market, product features, vertical integration level, economies of scale, geographical location, and competitive weapon to study the relationship between the management type of each hotel and business efficiency. The study's conceptual structure is seen in Fig. 9.1.

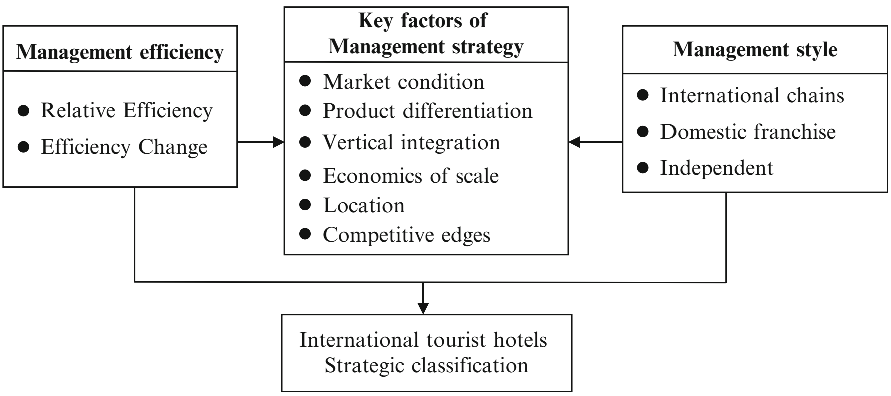


Fig. 9.1 Conceptual framework

9.2.2 Related Literature

Performance evaluation has been proven a vigor part of hotel management, not only be used as reference in decision-making, but also as the basis of improvement. Therefore, the measurement of efficiency becomes an important and broad-scope subject, particularly in hotel industry. For hotel operators, productivity is always a top priority (Brown and Dev 1999, 2000; Reynolds and Thompson 2007; Sigala 2004; Wang et al. 2006a, b). The concept of hotel productivity, like a spider web, wider includes efficiency, effectiveness, quality, predictability, and other performance dimensions (Sigala 2004). Brown and Dev (1999, 2000) also stated that productivity involves issues of efficient management, labor productivity (measurable), service productivity (elusive measures), and capital productivity.

As we can see, operational efficiency correlates closely to productivity. Therefore, efficient management becomes a main issue and objective for managers because this would affect the hotel productivity (Yang and Lu 2006; Brown and Ragsdale 2002; Sustainable Energy Ireland [SEI] 2001). In addition, many researches have proven that firms can improve hotel productivity through effective strategic decisions according to the demand and competitive conditions (Morey and Dittman 1995; Phillips 1996; Brown and Dev 1999).

However, there should be other factors coherent to productivity in hotel industry, as Anderson et al. (1999) suggested that the relative productivity of the hotel industry needed to take the mix and nature of services provided into account. Moreover, efficiency is often not easy to evaluate accurately because of difficultly determining an effective amount of resources. Recently, new techniques have been developed that have ability to compare the efficiency of similar service organization by explicitly considering their use of multiple inputs to produce outputs, including data envelopment analysis (DEA), the stochastic frontier approach, the thick frontier approach, and the distribution-free technique (Anderson et al. 1999).

Especially, DEA has been wider accepted and applied in many fields, such as in literature (Charnes et al. 1978), or in the performance/productivity of a hotel (Reynolds 2004; Reynolds and Thompson 2007). Wang et al. (2006a, b) employed the DEA and used the Tobit regression model to evaluate the efficiency determinants of the firms. Due to these researches, we could find out that DEA is an exceptional tool because it does not require an assumption about functional form of the model that underpins the relationships between the input and output variables (Hwang and Chang 2003).

In addition, Tsai et al. (2009) also pointed out that DEA is a rigorous tool in the analysis of firm-specific competitiveness factors because it provides a direct assessment between efficiency and financial performances. It also takes into consideration multiple input and output measurements to estimate relative efficiencies in the amount of decision-making units in international hotel chains. Especially, with the growing of prominent hotel groups, the use of DEA will be applied for more complex productivity analysis to obtain compilation of detail and quality data in more complicated and competitive environments. For example, the increasing number of strategic alliances among the various segments of hospitality industry will intensify competition and also enhance the advantages of incumbent firms. Consequently, re-structuring or redistribution of resources in a business firm will inevitably influence the measurement of efficiency and productivity.

Few researchers have applied DEA to measure the efficiency change of hotels. Therefore, with the purpose of providing updated knowledge on theories, concepts, ideas and empirical studies on competitiveness in the context of tourism destinations and the hotel industry, this paper uses data envelopment analysis (DEA), developed by Charnes et al. (1978), and the Malmquist productivity index expressed by Färe et al. (1992), to measure the managerial efficiency of 51 hotels in 2010 and the efficiency change of 51 hotels from 2005 to 2010. Finally, based on the results of managerial efficiency and operational features of the hotel industry. This paper is to discuss the management strategies formulation of the international tourist hotel industry in Taiwan based on the efficiency evaluation.

9.3 Measuring Efficiencies of the International Tourist Hotels

9.3.1 Object of Study

The study based its sample on the international tourist hotels included in the 2010 Business Analysis Report of International Tourist Hotel Operations in the Taiwan Area printed in 2011. The first stage of the study analyzes the business efficiency of 51 international tourist hotels and the second stage conducted cross timeframe analysis from the years 2005 to 2010. Basic international tourist hotel data is show in Table 9.1.

9.3.2 Defining Input–Output Factors

Looking at the related local and foreign literature on international tourist hotel business performance, definition of input output may be different according to the purpose of the study and the model used. Johnson and Ball (1989) used the concept of productivity to measure hotel performance; input variables included labor, capital, raw materials, and energy, while output variables were occupancy rates and number of meals. Jeffrey and Hubbard (1994) and Chow et al. (1998) focused on occupancy performance; major influencing factors included number of foreign visitors, amount of marketing and promotional budget, and number of rooms for travelers. Tsaur and Tsai (1999) established the capital function of international tourist hotels; output variable being business income and input variables being number of rooms, labor costs, raw material expense, and utilities. Keh et al. (2006)

Table 9.1 The basic information of international tourist hotels in Taiwan

No	Hotel	Area	Room	Management style	Type of visitors	Type of hotels
H1	Grand Hotel Taipei	1	402	B	E	G
H2	Ambassador Hotel	1	432	C1	E	G
H3	Imperial Hotel	1	288	A	E	G
H4	Gloria Prince Hotel	1	220	A	E	G
H5	Emperor Hotel	1	97	A	E	G
H6	Riverview Taipei Hotel	1	201	A	E	G
H7	Caesar Park Hotel	1	388	C2	D	G
H8	Golden China Hotel	1	215	A	E	G
H9	San Want Hotel	1	268	C1	D	G
H10	Brother Hotel	1	250	A	D	G
H11	Santos Hotel	1	287	A	E	G
H12	The Landis Taipei Hotel	1	209	A	D	G
H13	United Hotel	1	243	A	D	G
H14	Sheraton Taipei Hotel	1	692	C1	D	G
H15	Royal Taipei Hotel	1	202	C1	D	G
H16	Howard Hotels Taipei	1	606	B	D	G
H17	Grand Hyatt Taipei	1	865	C1	D	G
H18	Regent Taipei	1	569	C2	D	G
H19	Sherwood Taipei	1	343	C3	D	G
H20	Shangri-La's Far Eastern Plaza Hotel Taipei	1	420	C2	D	G
H21	The Westin Taipei	1	288	C1	D	G
H22	Hotel Kingdom	2	457	A	E	G
H23	Hotel Holiday Garden	2	274	A	D	G
H24	Kaohsiung Ambassador Hotel	2	457	C1	D	G
H25	Grand Hi-Lai Hotel	2	436	A	D	G
H26	Howard Plaza Hotel Kaohsiung	2	283	B	D	G
H27	The Splendor Kaohsiung	2	592	B	D	G
H28	Han-Hsien International Hotel	2	311	B	D	G
H29	Hotel National	3	319	A	D	G
H30	Plaza International Hotel	3	226	A	E	G
H31	Evergreen Laurel Hotel (Taichung)	3	354	A	D	G
H32	Howard Prince Hotel Taichung	3	155	B	D	G
H33	Splendor Hotel Taichung	3	222	B	D	G
H34	Astar Hotel	4	168	A	E	G
H35	Marshal Hotel	4	270	A	E	F
H36	Parkview Hotel	4	343	A	E	F
H37	Hualien Farglory Hotel	4	381	A	D	F
H38	Landis Resort Yangmingshan	1	50	A	D	F

(continued)

Table 9.1 (continued)

No	Hotel	Area	Room	Management style	Type of visitors	Type of hotels
H39	Lalu Hotel	5	96	A	D	F
H40	Hibiscus Resort	5	201	B	E	F
H41	Grand Hotel Kaohsiung	2	107	B	D	F
H42	Caesar Park Hotel Kenting	5	254	C3	D	F
H43	Howard Beach Resort	5	418	B	D	F
H44	Hotel Royal Chihpen	5	183	C1	D	F
H45	Taoyuan Hotel.	6	390	C1	E	G
H46	Hotel Royal Hsinchu	6	208	B	D	G
H47	Hsinchu Ambassador Hotel	6	257	A	D	G
H48	Hotel Tainan	6	152	A	E	G
H49	Tayih Landis Hotel Tainan	6	315	A	D	G
H50	Evergreen Laurel Hotel (Tainan)	6	197	B	D	G
H51	Formosan Naruwan Hotel	6	276	A	E	F

Note: Area: 1, Taipei; 2, Kaohsiung; 3, Taichung; 4, Hualien; 5, Scenic Area; 6, Others
Management style: A, Independent; B, Domestic chains; C, International chains (C1, Franchise; C2, Management Contract; C3, Membership)
Type of visitors: D, FIT (Foreign Independence tour); E, Group (Group tour)
Type of hotels: G, City hotel; F, Resort hotel

presented a three-stage measurement model from a marketing standpoint. Among these, efficiency measurement used total business expenditure and number of rooms as input variables and marketing expenditure as output variable; performance measurement used marketing expenditure as input variable and room and F&B revenues as output variables and; productivity measurement used total expenditure and number of rooms as input variables and room and F&B revenues as output variables. Hsieh and Lin (2009) used room revenue, F&B revenue, number of employees, business expenditure, occupancy rate, size of F&B size efficiency as output and input indicators and conducted Network DEA to measure hotel industry performance. Aside from these studies, many scholars have compared the efficiency of chain- and independently- operated hotels, most of them found that hotels with chain operations are more efficient than independently operated ones (Hwang and Chang 2003; Chiang et al. 2004; Wang et al. 2006a, b; Perrigot et al. 2009; Yu and Lee 2009).

Input resources for tourist hotels management include input material, staff, capital and equipment. These resources produce tangible and intangible services through front office and back office operations (Yasin et al. 1996). There were two primary revenues for tourist hotels in Taiwan: accommodation and meals. These constitute more than 80 % of total revenues of hotels. Other revenues include revenues from laundry, lease of store space, night clubs, service fee, all of which do not exceed 20 % of total revenues.

Among input factors, numbers of employees are used to represent input manpower, total floor area of room numbers and dining department is used to represent capital investments of hotel and operating expenses are used to represent cost of input changes. In summary, indicators used by the Taiwan Tourism Bureau for input–output factors are as follows:

Output Factors

- Room Revenue: refers to revenues from lease of rooms.
- Food and Beverages Revenue: refers to income derived from sale of food, snacks, alcohols, beverages in dining room, coffee room, banquet and night clubs.
- Other Revenues: refers to revenues other than the two items mentioned above. It includes operating revenues from lease of store spaces, laundry, swimming pool, ball courts, barber-shop, beauty salons and bookstores.

Input Factors

- Number of full-time Employees: refers to hired employees.
- Guest Rooms: refers to number of guest rooms in the hotel.
- Total Area of Meal Department: measured by total floor area.
- Operating Expenses: including salary, cost of meals, utility, fuel, insurance and other relevant operating costs.

9.3.3 Relative Efficiency Evaluation of Models

This paper uses data envelopment analysis (DEA), developed by Charnes et al. (1978). According to CCR model, developed by Charnes et al. (1978), the efficiency g_0 of a decision making unit (DMU) j_0 can be obtained by solving the following output- oriented CCR model:

$$\begin{aligned} 1/\mathrm{g}_0 = Min \ & \sum_{i=1}^{m} v_i x_{ij_0} \\ s.t. \ & \sum_{j=1}^{n} u_r y_{rj_0} = 1 \\ & \sum_{j=1}^{n} u_r y_{rj} - \sum_{j=1}^{n} v_i x_{ij} \leq 0 \\ & u_r, v_i \geqq \varepsilon > 0, \quad \mathrm{i} = 1, \ldots, \mathrm{m} \quad \mathrm{r} = 1, \ldots, \mathrm{s} \quad \mathrm{j} = 1, \ldots, \mathrm{n} \end{aligned} \tag{9.1}$$

For computational convenience, the efficiency of any DMU j_0 can be solved by dual of M1. The dual of M1 can be written as follows:

$$
\begin{aligned}
1/\mathrm{g}_0 = \text{Max} \ \ & \theta + \varepsilon\left(\sum_{i=1}^{m} s_{ij_0}^{-} + \sum_{r=1}^{s} s_{rj_0}^{+}\right) \\
s.t. \quad & \sum_{j=1}^{n} \lambda_j x_{ij} + s_{ij_0}^{-} = x_{ij_0} \\
& \sum_{j=1}^{n} \lambda_j y_{rj} - \theta y_{rj_0} - s_{rj_0}^{+} = 0 \\
& \lambda_j, s_{ij_0}^{-}, s_{rj_0}^{+} \geq 0, \quad \mathrm{i} = 1, \ldots, \mathrm{m} \\
& \mathrm{r} = 1, \ldots, \mathrm{s}, \quad \mathrm{j} = 1, \ldots, \mathrm{n} \\
& \theta \text{ unconstrained}
\end{aligned} \tag{9.2}
$$

The necessary and sufficient conditions for any DMU j_0 to reach efficiency are $\mathrm{g}_0 = \theta^* = 1, s_{ij_0}^{-*} = s_{rj_0}^{+*} = 0$, where a star superscript to a variable is used to denote its optimal solution (Charnes et al. 1978). For efficient DMUs, their efficiency value is 1 form the efficient frontier. The target provides benchmark for an inefficient DMU j_0 can be derived from $x_{ij_0}^{'} = x_{ij_0} - s_{ij_0}^{-*}$ and $y_{rj}^{'} = \theta^* y_{rj_0} + s_{rj_0}^{+*}$, where the slacks $s_{ij_0}^{-*}$ imply inputs surpluses and slacks $s_{rj_0}^{+*}$ imply outputs shortfalls. Besides, (9.2) identifies a set of corresponding efficient DMUs said to form a peer group for each inefficient DMU. Peer units are associated with basic λ_j. Since the efficiency value is 1 for all efficient DMUs on the efficient frontier, Anderson and Petersen (1993) proposed a modified model of (9.2) to increase the discrimination power for every efficient DMU by adding a constrain $j \neq j_0$.

9.3.4 Measurement of Efficiency Change

The method for measuring an organization's efficiency can be extended to measure the change of an organization's efficiency with the combination of the Malmquist productivity approach (Cave et al. 1982). As shown in Fig. 9.2, $\mathrm{F_t}$ represents the efficient frontier at period t, and $\mathrm{F_{t+1}}$ the efficient frontier at period $\mathrm{t+1}$. $A_t(x_t, y_t)$ and $A_{t+1}\left(x_{t+1}, y_{t+1}\right)$ represent the inputs-outputs vector of a DMU A at period t and $\mathrm{t+1}$, respectively. To propose the method for measuring the efficiency change from the time periods t to $\mathrm{t+1}$, the efficiency distance functions $D^{t+1}(x^t, y^t)$ are defined (which use the efficient frontier period $\mathrm{t+1}$ as the reference set for measuring the efficiency of a certain DMU A at period t, as the following linear programming problem:

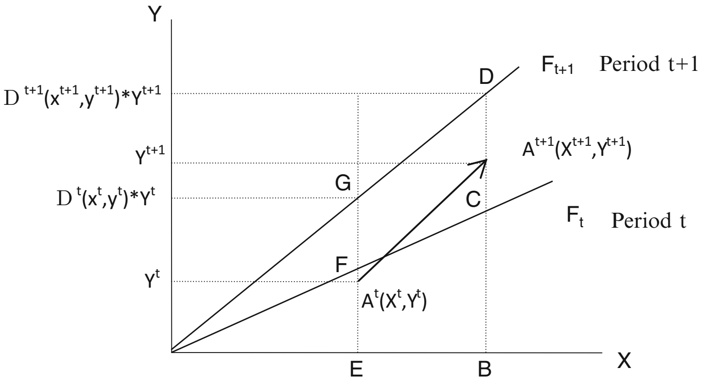


Fig. 9.2 The output based measurement of efficiency change

$$
\begin{aligned}
D^{t+1}(x^t, y^t) = \text{Max } \theta & \\
s.t. \quad & \sum_{j=1}^{n} \lambda_j^{t+1} x_{ij}^{t+1} \le x_{ij}^t \\
& \sum_{j=1}^{n} \lambda_j^{t+1} y_{rj}^{t+1} \ge \theta y_{rj0}^t \\
& \lambda_j^{t+1} \ge 0, \qquad \text{i} = 1, \ldots, \text{m} \\
& \text{r} = 1, \ldots, \text{s}, \quad \text{j} = 1, \ldots, \text{n} \\
& \theta \ \text{ unconstrained}
\end{aligned}
\tag{9.3}
$$

Similarly, we also can define $D^t(x^{t+1}, y^{t+1})$, which is to use the efficient frontier period t as the reference set for measuring the efficiency of a certain DMU A at period t + 1, as the following linear programming problem:

$$
\begin{aligned}
D^{t}(x^{t+1}, y^{t+1}) = \text{Max}\, \theta & \\
s.t. \quad & \sum_{j=1}^{n} \lambda_j^{t} x_{ij}^{t} \le x_{ij}^{t+1} \\
& \sum_{j=1}^{n} \lambda_j^{t} y_{rj}^{t} \ge \theta y_{rj}^{t+1}{}_0 \\
& \lambda_j^{t} \ge 0, \qquad \text{i} = 1, \ldots, \text{m} \\
& \text{r} = 1, \ldots, \text{s}, \quad \text{j} = 1, \ldots, \text{n} \\
& \theta \ \text{ unconstrained}
\end{aligned}
\tag{9.4}
$$

Obviously, both $D^t(x^t, y^t)$ and $D^{t+1}(x^{t+1}, y^{t+1})$ are output-oriented CCR model as (9.2). From the geometric meaning of aforementioned distance function in Fig. 9.1, we know that:

$$D^t(x^t, y^t) = EF/EA^t, \quad D^{t+1}(x^{t+1}, y^{t+1}) = BD/BA^{t+1},$$
$$D^t(x^{t+1}, y^{t+1}) = BC/BA^{t+1}, \quad D^{t+1}(x^t, y^t) = EA^t/EG$$

According to the Malmquist productivity index expressed by Färe et al. (1992) following Cave et al. (1982), the shift in efficiency (SIE) from period t to period t + 1 can be described by *BD*/*BC* and *EG*/*EF*. The geometric average of *BD*/*BC* and *EG*/*EF* can be used to measure the SIT, as represented by (M5)

$$SIE_{t,t+1} = \left[\frac{BD}{BC}\frac{EG}{EF}\right]^{1/2} = \sqrt{\frac{D^{t+1}(x^{t+1}, y^{t+1})D^{t+1}(x^t, y^t)}{D^t(x^{t+1}, y^{t+1})D^t(x^t, y^t)}} \tag{9.5}$$

Also the catching-up in efficiency (CIE) from period t to period t + 1 can be represented by (M6), which represents the ratio between the relative efficiency of a DMU at period t + 1 against that at period t.

$$CIE_{t,t+1} = \frac{BA^{t+1}/BD}{EA^t/EF} = \left[\frac{D^{t+1}(x^{t+1}, y^{t+1})}{D^t(x^t, y^t)}\right]^{-1} = \frac{D^t(x^t, y^t)}{D^{t+1}(x^{t+1}, y^{t+1})} \tag{9.6}$$

$CIE_{t,t+1} \times SIE_{t,t+1}$ can be used to measure the total efficiency change (TEC) from the time period t to period t + 1; that is

$$\begin{aligned} TEC_{t,t+1} &= CIE_{t,t+1} \times SIE_{t,t+1} \\ &= \frac{D^t(x^t, y^t)}{D^{t+1}(x^{t+1}, y^{t+1})}\sqrt{\frac{D^{t+1}(x^{t+1}, y^{t+1})D^{t+1}(x^t, y^t)}{D^t(x^{t+1}, y^{t+1})D^t(x^t, y^t)}} \\ &= \sqrt{\frac{D^t(x^t, y^t)D^{t+1}(x^t, y^t)}{D^t(x^{t+1}, y^{t+1})D^{t+1}(x^{t+1}, y^{t+1})}} \end{aligned} \tag{9.7}$$

(9.7) is the same as Malmquist productivity index, that is we use Malmquist productivity index as a measure for efficiency change (Fig. 9.2).

9.4 Managerial Efficiency of International Tourist Hotels

9.4.1 Relative Managerial Efficiency

Based on (9.2) CCR model and the Andersen and Petersen model, an evaluation of input–output information published in the Taiwan Tourism Ministry "Analytical Report on Management of International Tourist Hotels" was conducted in 2010 (Taiwan Tourism Bureau 2011). Results, in order of relative managerial efficiency, relative efficiency, reference set and frequency which DMU is in reference set, are shown in Table 9.2. Hotels with the value of 1 means that relative efficiency of hotels

Table 9.2 Efficiency analysis for international tourist hotels in 2010

No	Hotel	CCR efficiency	Reference groups	A&P efficiency	Frequency	Rank
H1	Grand Hotel Taipei	1.0000	H1	9.8003	1	1
H2	Ambassador Hotel	0.9843	H18, H48, H50	0.9843	0	15
H3	Imperial Hotel	0.6235	H7, H18, H50	0.6235	0	48
H4	Gloria Prince Hotel	0.7513	H7, H18, H42	0.7513	0	45
H5	Emperor Hotel	0.8072	H13, H34, H42	0.8072	0	39
H6	Riverview Taipei Hotel	0.9275	H7, H18, H50	0.9275	0	18
H7	Caesar Park Hotel	1.0000	H7	5.8410	25	1
H8	Golden China Hotel	0.8837	H7, H18, H50	0.8837	0	22
H9	San Want Hotel	0.8488	H18, H19, H20	0.8488	0	28
H10	Brother Hotel	0.8907	H18, H50	0.8907	0	21
H11	Santos Hotel	0.8375	H7, H18, H42	0.8375	0	31
H12	The Landis Taipei Hotel	0.8456	H18, H19, H42	0.8456	0	29
H13	United Hotel	1.0000	H13	2.4811	1	1
H14	Sheraton Taipei Hotel	0.9138	H7, H17, H18, H39, H42	0.9138	0	19
H15	Royal Taipei Hotel	0.9110	H18, H19, H20, H42	0.9110	0	20
H16	Howard Hotels Taipei	0.8106	H7, H18, H28, H39	0.8106	0	37
H17	Grand Hyatt Taipei	1.0000	H17	1.7664	1	1
H18	Regent Taipei	1.0000	H18	3.1782	28	1
H19	Sherwood Taipei	1.0000	H19	8.1835	4	1
H20	Shangri-La’s Far Eastern Plaza Hotel Taipei	1.0000	H20	5.23814	3	1
H21	The Westin Taipei	0.9389	H18, H39	0.9389	0	16
H22	Hotel Kingdom	0.8555	H7, H18, H42	0.8555	0	26
H23	Hotel Holiday Garden	0.8581	H7, H34, H42	0.8581	0	25
H24	Kaohsiung Ambassador Hotel	0.7722	H7, H18, H50	0.7722	0	44
H25	Grand Hi-Lai Hotel	0.8138	H18	0.8138	0	36
H26	Howard Plaza Hotel Kaohsiung	0.7936	H7, H34, H50	0.7936	0	41
H27	The Splendor Kaohsiung	0.6934	H7, H18, H50	0.6934	0	46
H28	Han-Hsien International Hotel	1.0000	H28	4.0805	5	1
H29	Hotel National	0.8716	H18, H28, H50	0.8716	0	23

(continued)

Table 9.2 (continued)

No	Hotel	CCR efficiency	Reference groups	A&P efficiency	Frequency	Rank
H30	Plaza International Hotel	0.7825	H7, H18, H50	0.7825	0	43
H31	Evergreen Laurel Hotel (Taichung)	0.8161	H7, H18, H50	0.8161	0	34
H32	Howard Prince Hotel Taichung	0.8515	H7, H18, H28, H50	0.8515	0	27
H33	Splendor Hotel Taichung	0.5476	H18, H48, H50	0.5476	0	49
H34	Astar Hotel	1.0000	H34	2.0881	9	1
H35	Marshal Hotel	0.8059	H7, H34, H50	0.8059	0	40
H36	Parkview Hotel	0.8318	H7, H18, H50	0.8318	0	32
H37	Hualien Farglory Hotel	0.8159	H7, H34, H42	0.8159	0	35
H38	Landis Resort Yangmingshan	1.0000	H38	21.4328	0	1
H39	Lalu Hotel	1.0000	H39	11.9023	3	1
H40	Hibiscus Resort	0.5420	H7, H34, H50	0.5420	0	50
H41	Grand Hotel Kaohsiung	0.4920	H18	0.4920	0	51
H42	Caesar Park Hotel Kenting	1.0000	H42	5.7204	13	1
H43	Howard Beach Resort	0.8182	H7, H34, H42	0.8182	0	33
H44	Hotel Royal Chihpen	0.8106	H19, H20, H42	0.8106	0	38
H45	Taoyuan Hotel.	0.8608	H7, H34, H42	0.8608	0	24
H46	Hotel Royal Hsinchu	0.9362	H7, H18, H50	0.9362	0	17
H47	Hsinchu Ambassador Hotel	0.6410	H7, H18, H28, H50	0.6410	0	47
H48	Hotel Tainan	1.0000	H18, H50	2.0119	1	1
H49	Tayih Landis Hotel Tainan	0.7897	H7, H18, H28, H50	0.7897	0	42
H50	Evergreen Laurel Hotel (Tainan)	1.0000	H50	1.6502	20	1
H51	Formosan Naruwan Hotel	0.8391	H7, H34, H42	0.8391	0	30

are all located at the efficient frontier of all 51 hotels. Reference groups of hotels with the value of less than 1 are hotels with the most relative efficiency, thus are all located in efficient frontiers. For example, the reference groups of Ambassador Hotel include Regent Hotel, Hotel Tainan and Evergreen Plaza Hotel (Tainan).

The results showed that there were 14 hotels with value of 1, namely The Grand Hotel, Caesar Park Taipei, United Hotel, Grand Hyatt Hotel, Regent Hotel, Sherwood Hotel, Far Eastern Plaza Hotel, Han-Hsien Hotel, Astar Hotel, Landis Resort, Lalu Hotel, Caesar Park Hotel Kenting, Hotel Tainan and Evergreen Plaza hotel. Some excel in output, while some have appropriate management of input resources. Sherwood Hotel, Regent Taipei, Grand Hyatt Hotel and Imperial hotel are part of an international alliance reservation organization and have excellent performance in terms of room revenues. The Meals revenues of Grand Hotel Taipei, Sherwood Hotel and Far Eastern Plaza Hotel are higher than room revenues. It can be said that their efficiencies are outstanding. Meanwhile, The Lalu Hotel and Caesar Park Hotel Taipei have efficiently utilized input resources such as human resources and equipment. Thus, they were all located at the efficient frontier.

Hotels which have relatively poor efficiency values include Splendor Hotel Taichung, Tsengwen Hibiscus Resort and Grand Hotel Kaohsiung, with a value of less than 0.6. Inasmuch as room revenues, meals revenues and other revenues are less than other hotels, their efficiency for the current period is relatively poor. The Splendor Kaohsiung invested an excessively high amount on operating expenses and has poorly utilize its human resources which explains its poor efficiency. Tsengwen Hibiscus Resort's output efficiency indicator performance is far lower than the standards of industrial peers and revenues from rooms and meals are not ideal. The reason why Grand Hotel Kaohsiung performs poorly is due to the fact that it has a poor room efficiency.

Generally speaking, the overall results show an average efficiency value of 0.8591, a standard deviation of 0.1292, there were 14 hotels with an efficiency value of 1.

9.4.2 Change of Managerial Efficiency

The model for measuring the efficiency change, as mentioned previously in this paper, is used to examine the managerial efficiency change of international tourist hotels over the year from 2005 to 2010. The results are listed in Table 9.3 in order of the value of efficiency change.

The results indicate there were 9 hotels with efficiency change gather than 1. This means that over the past 4 years, managerial efficiency of 9 hotels has been improving, with Evergreen leading. There were 42 hotels with efficiency change less than 1. This means that managerial efficiency change of 42 hotels has been declining with splendor Hotel Taichung has poorest record. This is because suffered from financial crisis in 2008, which is influenced inflation.

The correlation coefficients between efficiency change and following items: relative efficiency in 2005, relative efficiency in 2010, Shift in technology in 2005–2010 and catch up efficiency in 2005 to 2010 are listed in Table 9.4. It showed that the efficiency change of 51 international hotels between 2005 and 2010 in Taiwan is positively correlated with relative efficiency in 2010. However, the efficiency change is negatively correlated with relative efficiency in 2005. For example, H50, H51 and H22 have lagging relative efficiency in 2005 but made great improvement over the past 5 years. H4, H5 and H39 have a good relative efficiency in 2005 but little efficiency change in the past 5 years.

Table 9.3 Efficiency change analysis for tourism hotels in 2005–2010

No	Hotel	CIE	SIT	EC	No	Hotel	CIE	SIT	EC
H1	Grand Hotel Taipei	1.7190	0.6974	1.1988	H27	The Splendor Kaohsiung	1.0194	0.8653	0.8821
H2	Ambassador Hotel	0.8228	0.8065	0.6636	H28	Han-Hsien International Hotel	1.5718	0.5645	0.8873
H3	Imperial Hotel	0.7744	0.8674	0.6717	H29	Hotel National	0.9896	0.7079	0.7006
H4	Gloria Prince Hotel	0.5998	0.6269	0.3760	H30	Plaza International Hotel	0.7104	0.7787	0.5532
H5	Emperor Hotel	0.4099	0.9005	0.3691	H31	Evergreen Laurel Hotel (Taichung)	0.8220	0.8795	0.7230
H6	Riverview Taipei Hotel	1.1420	0.9115	1.0409	H32	Howard Prince Hotel Taichung	0.9137	0.8759	0.8004
H7	Caesar Park Hotel	1.0387	0.9618	0.9990	H33	Splendor Hotel Taichung	0.6040	0.5984	0.3614
H8	Golden China Hotel	0.9336	0.8459	0.7897	H34	Astar Hotel	1.0059	1.0003	1.0061
H9	San Want Hotel	1.0552	0.7060	0.7450	H35	Marshal Hotel	0.9335	0.7078	0.6608
H10	Brother Hotel	0.7028	0.8495	0.5971	H36	Parkview Hotel	0.9178	0.7858	0.7212
H11	Santos Hotel	1.0465	0.9618	1.0065	H37	Hualien Farglory Hotel	0.6189	0.9446	0.5846
H12	The Landis Taipei Hotel	0.8830	0.7483	0.6607	H38	Landis Resort Yangmingshan	1.9365	0.4311	0.8348
H13	United Hotel	0.9829	0.9072	0.8917	H39	Lalu Hotel	0.7473	0.5990	0.4476
H14	Sheraton Taipei Hotel	1.3222	0.6923	0.9154	H40	Hibiscus Resort	0.8270	0.8554	0.7074
H15	Royal Taipei Hotel	1.0842	0.7663	0.8309	H41	Grand Hotel Kaohsiung	0.6535	0.9319	0.6090
H16	Howard Hotels Taipei	0.8331	0.7250	0.6041	H42	Caesar Park Hotel Kenting	1.2906	0.7751	1.0003
H17	Grand Hyatt Taipei	0.9328	0.8829	0.8235	H43	Howard Beach Resort	0.5554	1.0659	0.5920
H18	Regent Taipei	1.0979	0.8164	0.8963	H44	Hotel Royal Chihpen	0.7545	0.7308	0.5513
H19	Sherwood Taipei	0.9599	0.6774	0.6502	H45	Taoyuan Hotel.	0.6998	0.9151	0.6404
H20	Shangri-La's Far Eastern Plaza Hotel Taipei	1.2076	0.8790	1.0615	H46	Hotel Royal Hsinchu	1.1040	0.8810	0.9726
H21	The Westin Taipei	0.8008	1.0303	0.8251	H47	Hsinchu Ambassador Hotel	0.7192	0.7871	0.5660
H22	Hotel Kingdom	1.2499	0.8551	1.0688	H48	Hotel Tainan	0.9665	0.7835	0.7573
H23	Hotel Holiday Garden	0.9302	0.7983	0.7426	H49	Tayih Landis Hotel Tainan	0.9987	0.8728	0.8717
H24	Kaohsiung Ambassador Hotel	0.8417	0.7040	0.5925	H50	Evergreen Laurel Hotel (Tainan)	1.9343	0.8800	1.7022
H25	Grand Hi-Lai Hotel	0.9933	0.8885	0.8825	H51	Formosan Naruwan Hotel	1.0981	0.9817	1.0781
H26	Howard Plaza Hotel Kaohsiung	0.9634	0.8843	0.8519					

Table 9.4 Correlation coefficients between efficiency change and efficiency of the tourism hotels in Taiwan

Efficiency	Correlation coefficients with efficiency change
Relative efficiency in 2005	−0.3606***
Relative efficiency in 2010	0.4811***
Shift in technology (SIT) in 2005–2010	0.8056***
Catching-up in efficiency (CIE) in 2005–2010	0.2775**

9.5 The Relationship Between Managerial Performance and Management Strategy

9.5.1 The Variance Analysis of Hotel Characteristics and Managerial Performance

This study further classified tourist hotels in accordance with different market conditions, location, sources of visitors, room size, management style and ranking to study the difference on managerial efficiency of different types of hotel as Table 9.5. Result of nonparametric Statistics reveled there were no significance in managerial efficiency due to source of visitors, area and market condition. However, there were significant differences in managerial efficiency due to differences in management style. As a result, evaluating a managerial efficiency of international hotel in Taiwan, is related its internationalization. For improving managerial efficiency, it is crucial to strength management techniques, to initiate competitive advantages and performance.

The management style can be divided in to three categories, international chains, domestic chains and independent. The management of efficiency of Domestic chain hotels was much less than international chain hotels. Of the three operating methods for international chains, efficiency value is highest on hotels which commissioned a professional hotel group to manage a hotel on its behalf through means of a management contract.

9.5.2 Strategic Factor of the Hotel Industry

Through analysis of strategic elements, we can understand the possible business strategy and crucial characteristics of the hotel industry, which can be further categorized into appropriate strategy activities. Through the results of the abovementioned variance test, the performance of international tourist hotels vary due to a difference in management type, including international cooperation, local chain, and independent operation. Six value-added activities considered as factors behind the strategies and decisions of the hotel industry including target market, product features, vertical integration level, economies of scale, geographical location, and competitive weapon are also further explained (see Table 9.6).

Table 9.5 Difference verification of the efficiency for the 51 tourism hotels in Taiwan

Classification		Sample size	Efficiency	Std. dev.	Mann-Whitney U (*p*-value)	Kruskal-Wallis χ^2 (*p*-value)
Market condition	City hotels	39	0.8681	0.1155	−0.606 (0.545)	0.367 (0.545)
	Resort hotels	12	0.8296	0.1689		
Source of visitors	FIT	34	0.8671	0.1316	−0.878 (0.380)	0.771 (0.380)
	Group	17	0.8431	0.1268		
Location	Taipei	25	0.8964	0.1088	–	5.23 (0.156)
	Taichung	7	0.7730	0.1700		
	Kaohsiung	11	0.8244	0.1507		
	Scenic area	8	0.8652	0.0839		
Room size	Under 200 rooms	9	0.8845	0.1707	–	3.124 (0.373)
	201–300 rooms	21	0.8192	0.1330		
	301–400 rooms	10	0.8841	0.0834		
	Above 401 rooms	11	0.8914	0.1129		
Management style	International chains	15	0.9317	0.0772	–	9.545 (0.08) **
	Domestic chains	13	0.7696	0.1763		
	Independent	23	0.8628	0.0940		

9.5.2.1 International Cooperation Management Style

Three ways are used for the international cooperation management style: joining an international chain hotel system, signing a management contract, and entrusting management to the international hotel chain or joining a reputable worldwide hotel organization to become a member hotel. Representative hotels include the Hyatt, Far Eastern Shangri-La, Sherwood, and the Ritz.

International chain hotels in the Taiwan area are located in Taipei City's commercial center. The target market is very clear, these hotels are primarily urban business hotels, serving mainly business or conference/exhibition travelers. Methods of travel include Foreign Individual Travelers (FIT) and Group. Because FIT is primarily made up of business travelers, they usually stay for longer periods of time and in rooms that are in the higher price range. Moreover, the sources for these travelers are stable and easier to handle, their travel is not affected seasonally and brand loyalty is higher. Consequently, hotels with higher percentages of FIT find it easier to create higher quality of service.

For hotels joining international cooperation, in the value-added activities of strategic elements, there are competitive advantages like the transfer of

Table 9.6 Managerial strategy classification of International tourist hotels

Strategic classification	International chains		Domestic chains		Independent	
	Large-size hotel	Small and medium	City hotel	Resort hotel	City hotel	Resort hotel
Representative	Grand Hyatt Taipei, Far eastern, Westin Taipei, Caesar Park Hotel Kenting	Sherwood Taipei, The Landis, Imperial Taipei	Howard, Evergreen hotel, Regent Taipei, Hotel Royal Taipei	Kenting Howard, Hotel Royal Chihpen, Fleur De Chine	United Hotel, Emperor Hotel, Miramar garden, Hilai Hotel	The Lalu, Parkview, Formosan Naruwan
Market condition	International-oriented	International-oriented	International-oriented	Asia/International-oriented	International/Domestic-oriented	Asia/Domestic-oriented
Product differentiation	Business Hotel FIT/Group	Business FIT	Business FIT/Group	Domestic FIT Japanese Group	Business FIT	Domestic FIT Japanese Group
Vertical integration	Diversification (Self-Operation / Outsourcing)	Partial integration (Self-Operation / Outsourcing)	Diversification and Inter-industry alliance	Diversification and Inter-industry alliance	Combining with upstream, and downstream inter-industry	Combining with upstream, and downstream inter-industry
Economics of scale	Global Reservation, Joint promotion, Management techniques, Personnel Training	Global Reservation, Joint promotion, Management techniques, Personnel Training	Joint Reservation, Joint promotion, Management techniques, Personnel Training	Joint Reservation, Joint promotion, Management techniques, Personnel Training	Inter-industry alliance, Joint promotion	Inter-industry alliance, Joint promotion, Developing new market segmentation in light season
Location	Decentralized management in Taipei City	Decentralized management in Taipei City	Metropolitan area, Decentralized management	Scenic area, Regional character	Metropolitan area, Regional character	Scenic area, Regional character
Competitive edges	International chains, Location, Product line stretching	Innovation ability, Location, International chains, Location	Diversification, Location	Diversification, Location	Location, price discrimination	Location

international management knowledge, the breadth and features of the product line, advantage of geographical locations, and economies of scale of joint room reservations. Hence, business performance is significantly better than hotels with other management styles.

9.5.2.2 Domestic Franchise Management Style

In recent years, amidst the competitive environment of international business management and accumulation of practical experience, local hoteliers have already learned to stay on top of the market pulse and business techniques. Many independently operated hotels have gradually made the transition to creating their own tourist hotel brands. These include the Howard Plaza, Evergreen, FIH Regent, Royal, and L'Hotel de Chine. Aside from joint marketing advantages such as joint room reservation and promotion as well as management technique sharing and manpower support to enhance image and visibility, local chain hotels do not have to deal with the membership conditions set by international chain systems. It also eliminates the limitations set by the enormous amount of management fee demanded by the foreign mother company as well as the difficulties in controlling and deploying manpower.

According to the target market and geographical characteristics, local chain hotels can be divided into two types: business hotels and leisure hotels. The product features of business hotels and the abovementioned business hotels of international cooperation are very similar. On the other hand, because they are mostly determined by the characteristics of geographic segmentation, the product features of leisure hotels should be based on the geographical features of the scenic area, the positioning of the target market, as well as the resources and features of the hotel per se, complemented by the use of product design and service strategies.

Summarizing the strategic elements of local chain hotels, the "competitive weapons" is "diversified product line extension strategy." This is using diversification business methods to develop domestic franchise hotels or the direction of related services, creating a synergy enhancing corporate brand image through exploiting advantageous conditions. Moreover, combining the strategy of cross-industry alliance of upstream and downstream tourism industries allows for complementary symbiosis, forming economies of scale and creating added value for travelers, resulting in competitive advantage.

9.5.2.3 Independent Management Style

Independent operation means the investor does not ask for outside assistance, preferring to be responsible for management and decision-making. Examples include Brother Hotel, Emperor Hotel, Tainan Hotel, and others. Based on "product features" and "geographical location," independently operated hotels can divided into two major types: business hotels located in urban areas in the central and southern parts of Taiwan as well as leisure hotels in scenic areas.

These hotels use vertical integration, working with airlines and the travel industry to promote domestic package tours to create added value for customers, thereby achieving the economies of scale of joint promotions. In terms of "geographical location," the business hotels are usually located in southern and central parts of Taiwan. The main customers are local business travelers and these hotels create geographical advantage by virtue of its location. The market is segmented geographically; local newspapers and travel agencies as well as other local media and marketing channels are used to reinforce working relationships with local companies and expand the customer base. For example, United Hotel, located in Taipei's commercial center, emphasizes its professional services meeting the needs of American and European travelers and uses mid-range pricing strategy to separate it from the other international business hotels. This strategy has attracted businessmen who prioritize functionality and affordability, resulting in good occupancy performance. Tainan Hotel and Grand Hi-Lai Hotel take advantage of geographical segmentation, competing for the international business travelers' market in Southern Taiwan.

Summarizing the various strategy elements, for the "competitive weapon" element of independently operated hotels, what is worth thinking about is how to exploit the advantages offered by the hotels' geographical location and available resources, avoiding direct conflict with international chain hotels, and using differentiated pricing strategy to attract different segmented target markets. Furthermore, adopting the cross-industry alliance strategy of vertically integrated upstream and downstream tourism industries is a good move on the part of independently operated hotel to reach economies of scale and create customer value. Hotels, which do not operate as chains, should use this comparative advantage strategy to find the difference between itself and its competitors in terms of competitive factors and exploit these factors fully to create comparative competitive advantage.

9.5.3 The Relationship Between Strategy Group and Managerial Performance of International Hotels

The values of relative efficiency for measuring the competitiveness are represented on the horizontal axis. A smaller value represents a hotel with less competitiveness and managerial efficiency, is much urgent for improvement. A large value represents a hotel with more competitiveness and managerial efficiency. In addition, measuring the competitiveness are represented on vertical axis. A smaller value represents that managerial efficiency with a large move, has a potential of continuous development, vice versa. Based on the demand for and rate of progress of improvement, the 51 hotels in Taiwan can be classified into five categories as follows:

9.5.3.1 Hotels with High Competitiveness and a Fast Pace of Progress (Category A in Fig. 9.3)

These include Evergreen Hotel (Tainan), Far East Hotel, the Grand Hotel, Astar Hotel, Caesar Park Hotel (Kenting). These hotels have a relative efficiency of above 1 and an efficiency change above 1.0. Currently these hotels have excellent managerial efficiency. Over the past 5 years, these hotels have improved rapidly. This signifies that they are on the right track.

These groups of hotels are mostly international or local or hotel chains or conferral or management contract or cooperative international ventures or a locally created brand; these are the hotels considered to have management efficiency. For large-scale international business hotels which cater mainly to foreign individual tourists (FIT), in terms of strategy elements, because they have more rooms, they should receive group travelers in addition to FIT to raise the overall occupancy performance. Consequently, planning and designing separate floors for business and group travelers with difference décor and flow of movement may contribute to the customer segmentation. Small-scale international hotels with fewer rooms may

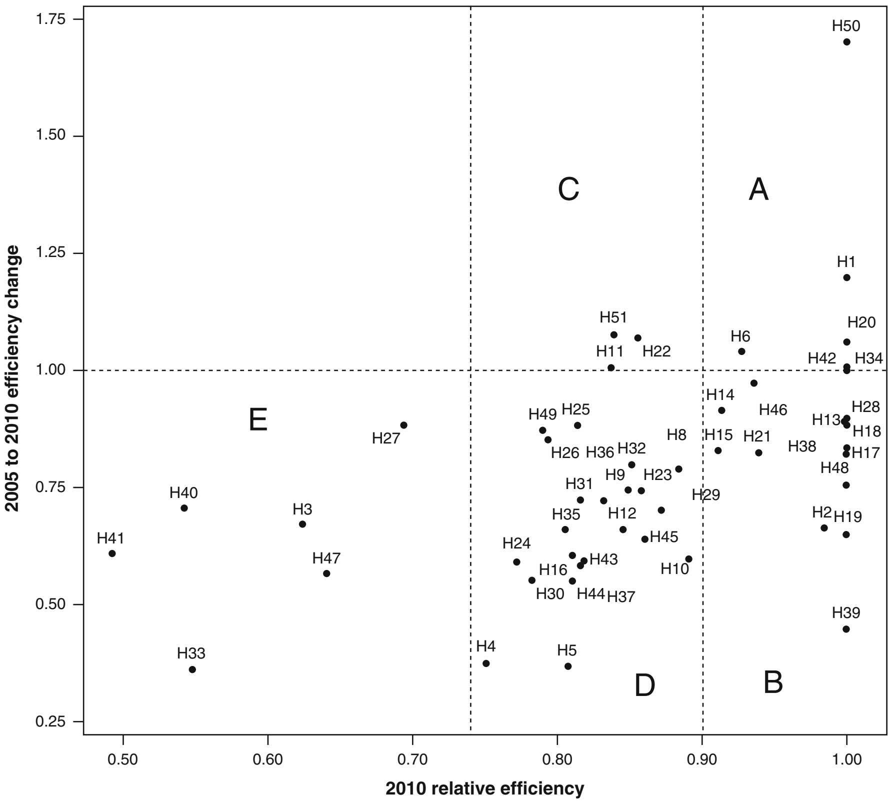


Fig. 9.3 The matrix of management decision for the 51 international hotels in Taiwan

use a differentiation strategy, emphasizing a more refined luxury. Hence, the hotels in this area may lock on to FIT as their target market, emphasizing the unique professional services specially designed for business travelers and refusing to receive group travelers. Features exclusive to this type of hotels will include room décor and flow of movement catering to business travelers, business center, international conference facilities, and innovative services.

In addition, in terms of the "economies of scale" element, international chain hotels have a purchasing scale of global room reservation, possessing a jointly shared reservation system, and it is the most advantageous way of achieving economies of scale. Looking at global competition, distributing and allocating the resources in the most appropriate manner and through international chain management (cooperation), it is possible to achieve the goals of technical transfer, joint room reservations, joint purchasing as well as personnel training and exchange. Joining the globally connected room reservation system is an important way for local hotels to acquire an international customer base. It is also one of the major reasons why many hotels choose to join a chain organization.

In summary, the positioning of the international chain business hotel market of this segment is very clear. It has the advantage of international hotel management and cooperation, which is complemented by competitive geographical location; resources are efficiently distributed, allowing for better business efficiency. An exemplary model for other hotels, this signifies that these hotels are doing the right thing business-wise. They should continue to be profitable and able to maintain market dominance.

9.5.3.2 Hotels with High Competitiveness but a Slower Pace of Progress (Category B in Fig. 9.3)

Grand Hyatt, Regent Taipei, Sheraton Taipei, Sherwood Taipei, Hotel royal Hsinchu, Hotel Royal Taipei, Ambassador Taipei, Landis Resort Yangmingshan, The Lalu Sun moon lake, Plaza International Hotel, United Hotel and The Westin Taipei, the change efficiency is less than 1, these hotels have excellent performance so far. However, it has lagging improvement in these 5 years.

Hotels in this area display good business performance during this timeframe, thus belonging to the relatively efficient group. These are mostly hotels with international or local chains; the rest are independently operated business hotels. In addition, these hotels use various marketing tools like forming strategic alliances with airlines and credit card companies as well as getting involved in charitable activities to expand their customer base, raise visibility, and enhance corporate image.

Aside from the abovementioned direction of making room revenue a main focus, furthering the role of F&B banqueting—developing banqueting needs of local consumers in the area is also a development focus. Furthermore, there is a need to improve the hotel's exteriors and brand of service; style of renovation needs to complement the unique features of the area. The strategy is for the hotel to become representative of the area it is located in.

For leisure hotels, they should find a way to integrate the resources of the local scene, developing exciting and creative travel design. Using the vertical integration model, they should form strategic alliances with airlines and travel agencies and step up low season promotion to improve occupancy and F&B performance. In doing this, the hotels should be able to become part of the highly efficient local chain hotel group.

9.5.3.3 Hotels with Medium Competitiveness but a Fast Pace of Progress (Category C in Fig. 9.3)

These include Hotel Kingdom, Santos Hotel, Formosan Naruwan Hotel. These Hotels have a relative efficiency of more than 0.85 but less than 0.9 and an efficiency change of more than 1. Currently, these hotels have medium managerial efficiency. However, they have experienced rapid efficiency change for the past 5 years. This means that the competitive advantage of these hotels is gradually increasing.

Hotels located in this area are primarily international chain hotels with plain performance; these include Rebar, Royal, and Ritz; main customers are foreign business travelers. Having the advantages from international cooperation, the hotels of this area should make adjustments in product features and resource distribution according to the needs of Japanese, European, or American business travelers. Becoming hotels distinct for their fine service and raising business efficiency are crucial strategic points. In addition, it is necessary to improve occupancy performance; using and promoting marketing strategies to reduce the difference in occupancy rates during peak and low seasons as well as to expand the F&B banqueting market by exploiting the advantage of being geographically accessible. In summary, the hotels in this area is a high potential group, which should be more aggressive in raising the level of efficiency to take the next step up towards being one of the highly efficient hotels.

9.5.3.4 Hotels with Medium Competitiveness but a Slow Pace of Progress (Category D in Fig. 9.3)

These include Tayih Landis, Golden China, San want, Grand Hi-Lai, Howard Hotels (Taichung, Kaohsiung, Kenting), Evergreen international (Taichung), Ambassador Taipei, Hotel Royal Chihpen, The Landis Taipei. These hotels have a relative efficiency of more than 0.75 but less than 0.9. Currently, these hotels have a medium managerial efficiency. Over the past 5 years, there has been a slight decline in efficiency change in these hotels, there is a need for these hotels to improve and catch up with other hotels.

The hotels in this area need to improve in terms of business efficiency. Their main customers are individual or group business travelers from the US, Europe, and parts of Asia. Management should be focused on improving improper resource distribution to control business efficiency. Independently operated hotels which do not belong to any international chain should use this comparative advantage strategy and look for the competitive difference between itself and its competitors as well as fully exploiting this difference to create comparative competitive advantage. First, the hotel must do geographic segmentation, taking advantage of the area features; its target customers and market positioning should be very clear for the hotel to be able to attract international business travelers to that specific area. For leisure hotels with local travelers as their target customers, they can attract the local travel customer group who like high standard and luxurious vacation quality. However, leisure hotels are frequently affected by the difference in low and peak seasons. Consequently, developing a low season market is one effective way to go; hotels may offer higher discount rates during low season to reach new customers. In addition, independently operated hotels should find a way to respond timely to any advantages offered by environmental changes and decision-making should be extremely flexible. They should also form cross-industry alliances with upstream and downstream companies using the vertical integration strategy to provide diverse value-added services, creating customer value and improving business performance.

9.5.3.5 Hotels with Low Competitiveness but a Slow Pace of Progress (Category E in Fig. 9.3)

These include Imperial Hotel, Ambassador Hotel (Hsinchu), The Splendor Kaohsiung, The Grand Hotel (Kaohsiung), Hibiscus Resort. These hotels have a relative efficiency of less than 0.7 and efficiency change between 0.6 and 0.9. Currently, the competitiveness of these hotels are clearly lagging behind others. Over the past 5 years, managerial efficiency has been declining.

The hotels in this area are independently operated but have poor business performance. The main target customers are individual travelers and secondary target customers are group travelers; primary customer base is domestic and Japanese. Study of the room and F&B performance finds that the common problem is the unprofitability of the room department. Because occupancy rate is low and the manpower input is in excess, it results in inefficient use of resources, which is the main reason for falling behind other hotels in terms of business efficiency. It is recommended that hotels in this area should aggressively look into the current distribution and management system, modifying their strategies in terms of the overall industry development and market positioning. They should also study their available resources and competencies to come up with short-term and long-term improvement steps and strategies (Fig. 9.3).

9.6 Conclusion

As weekend travel becomes more and more a priority for most people, many business groups are competing to gain foothold into the hotel industry, which also means stiffer competition for those already in the industry. To create a stronger competitive advantage, it is first necessary to understand the business performance of the company in the said industry to enhance its competitiveness. This study makes use of the DEA model of Charnes et al. (1978) in evaluating the relative business performance of Taiwan's tourist hotels. At the same time, this is combined with the Malmquist Productivity Index from Färe et al. (1992), which evaluates the growth and reduction situation of management performance of 51 tourist hotels in Taiwan from 2005 to 2010.

Study results show that, overall, Taiwan's hotel industry show slight technical change; overall technical standards have improved, but not significantly. Efficiency change is higher than "catch-up in efficiency," meaning that production technology within the industry has improved and that it has resulted in the outward movement of the production frontier. Moreover, analysis of relative efficiency and technical change management matrix found a dire need for long-term planning controls to prevent worsening of management efficiency.

In addition, the study also looks into the effects of differences in business style, geographical location, number of guest rooms, and management style on business efficiency. It found that management efficiency of Taiwan's international hotels is significantly different due to management style, while there are no significant differences in terms of the nature of the hotel and geographical location. It is obvious from this that, to improve the management efficiency of Taiwan's international hotels, it is necessary to strengthen international management techniques, create competitive advantage, and produce business performance. Consequently, the study looks at the differences in management style and management efficiency and forms appropriate management strategies based on the results of the strategic value chain analysis.

Finally, based on the results of business efficiency gleaned from the situation of the total efficiency of the specific timeframe and the changes in technical efficiency across timeframes, the study divides the efficiency performance of international hotels into five groups in terms of three aspects: management style, strategic elements, and business performance to study the characteristics and business strategy of each group. Results of the study show that joining an international hotel chain is an important way for a hotel to gain unique skills. However, what is worth noting is, technical or management skills gained from franchising or affiliation are more of "resources" rather than of "skills." Only when these resources can be translated and internalized into organizational competencies, becoming a part of the operation of the organization can it cope with changes within the environment and maintain long-term advantage.

Aside from using data from a specific timeframe to analyze and measure the business performance of the international tourist hotel industry, the study systematically uses the DEA model to also conduct passive analysis on changes in technical efficiency across various timeframes. Combined with the results from the evaluation of relative efficiency and efficiency change, this could provide a management strategy matrix, which is a valuable reference for managers in the hotel industry. It could also be an appropriate evaluation tool for supervising and auditing work in the hotel business.

References

Anderson P, Petersen NC (1993) A procedure for ranking efficiency units in data envelopment analysis. Manag Sci 39:1261–1264

Anderson RI, Fish M, Xia Y, Michello F (1999) Measuring efficiency in the hotel industry: a stochastic frontier approach. Int J Hosp Manag 18(1):45–57

Ansoff HI (1990) Implanting strategic management, 2nd edn. Prentice Hall, Englewood Cliffs

Brown JR, Dev CS (1999) Looking beyond RevPAR: productivity consequences of hotel strategies. Cornell Hotel Restaur Admin Q 40(2):23–33

Brown JR, Dev CS (2000) Improving productivity in a service business: evidence from the hotel industry. J Serv Res 2(4):339–354

Brown JR, Ragsdale CT (2002) The competitive market efficiency of hotel brands: an application of data envelopment analysis. J HospTour Res 26(4):332–360

Cave DW, Christensen LR, Diewert WE (1982) The economic theory of index numbers and the measurement of input, output, and productivity. Econometrica 50:1393–1414

Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision making units. Eur J Oper Res 2(6):429–444

Chiang WE, Tsai MH, Wang LSM (2004) A DEA evaluation of Taipei hotels. Ann Tour Res 31 (3):712–715

Chow WS, Shyu JC, Wang KC (1998) Developing a forecast system for hotel occupancy rate using integrated ARIMA models. J Int Hosp Leis Tour Manag 1(3):55–80

Conner KR (1991) A historical comparison of resource-based theory and five schools of thought within industrial organization economics: do we have a new theory of the Firm? J Manag 17 (1):121–154

Day GS (1994) The capabilities of market driven organization. J Int Bus Stud 11(1):37–52

Färe R, Grosskopf S, Lindgren B, Roos P (1992) Productivity changes in Swedish pharmacies 1980–1989: a non-parametric malmquist approach. J Prod Anal 3(1):85–101

Hsieh LF, Lin LH (2009) A performance evaluation model for international tourist hotels in Taiwan—an application of the relational network DEA. Int J Hosp Manag 29:14–24

Hwang SN, Chang TY (2003) Using data envelopment analysis to measure hotel managerial efficiency change in Taiwan. Tour Manag 24(4):357–369

Jeffrey D, Hubbard NJ (1994) A model of hotel occupancy performance for monitoring and marketing in the hotel industry. Int J Hosp Manag 13(1):57–71

Johnson K, Ball S (1989) Productivity measurement in hotels. Tourism marketing and management handbook. Prentice-Hall, Englewood Cliffs, pp 319–323

Keh HT, Chu S, Xu J (2006) Efficiency, effectiveness and productivity of marketing in services. Eur J Oper Res 170:265–276

Miller D, Snow CC (1978) Organization strategy, structure and process. McGraw-Hill, New York

Morey RC, Dittman DA (1995) Evaluating a hotel GM's performance: a case study in benchmarking. Cornell Hotel Restaur Admin Q 36(5):30–35

Okavarrueta S (1996) Market attractiveness, resourced-based and evolutionary approaches to strategy: a comparison. In: Wilson EJ, Hair JF (eds), Developments in marketing science: Proceedings of the 1996 Academy of Marketing Science. Springer, pp 149–153

Perrigot R, Cliquet G, Piot-Lepetit I (2009) Plural form chain and efficiency: insights from the French hotel chains and the DEA methodology. Eur Manag J 27:268–280

Phillips PA (1996) Strategic planning and business performance in the quoted UK hotel sector: results of an exploratory study. Int J Hosp Manag 15(4):347–362

Porter ME (1980) Competitive strategies. Free Press, New York

Porter ME (1991) Towards a dynamic theory of strategy. Strateg Manag J 12(1):95–117

Prahalad CK, Hamel G (1990) The core competence of the corporation. Harv Bus Rev May–June:277–299

Reynolds D, Thompson GM (2007) Multiunit restaurant productivity assessment using three-phase data envelopment analysis. Int J Hosp Manag 26(1):20–32

Reynolds D (2004) An exploratory investigation of multiunit restaurant productivity assessment using data envelopment analysis. J Travel Tour Mark 16(2/3):19–26

Roth AV, Jackson WE, III (1995) Strategic determinants of service quality and performance: evidence from the banking industry. Manag Sci 41(11):1720–1733

Sigala M (2004) Using data envelopment analysis for measuring and benchmarking productivity in the hotel sector. J Travel Tour Mark 16(2/3):39–60

Sustainable Energy Ireland (2001) Managing energy: a strategic guide for hotels. http://www.sei.ie. Accessed 10 May 2007

Taiwan Tourism Bureau (2005–2010) The operating report of international tourist hotels in Taiwan. Republic of China Press, Taipei

Tsai H, Song H, Wong KF (2009) Tourism and hotel competitiveness research. J Travel Tour Market 26(5):522–546

Tsaur SH, Tsai CW (1999) Cost structure for international tourist hotels in Taiwan. Asia Pac J Tour Res 4(1):53–64

Wang FC, Shang JK, Hung WT (2006a) Productivity and service quality changes in international hotels in Taiwan. Ann Tour Res 33(2):571–574

Wang FC, Hung WT, Shang JK (2006b) Measuring the cost efficiency of international tourist hotels in Taiwan. Tour Econ 12(1):65–85

Yang C, Lu WM (2006) Performance benchmarking for Taiwan's international tourist hotels. INFOR 44(3):229–245

Yasin MM, Andrew JC, Jeffrey JD (1996) A framework for the establishment of an optimal service quality level in a hospitality operational setting. J Hosp Leis Market 4(2):25–48

Yu MM, Lee YC (2009) Efficiency and effectiveness of service business: evidence from international tourist hotels in Taiwan. Tour Manag 30:571–580

Chapter 10
Sustainable Product Design Performance Evaluation with Two-Stage Network Data Envelopment Analysis

Chialin Chen, Joe Zhu, Jiun-Yu Yu, and Hamid Noori

Abstract Data Envelopment Analysis (DEA) has traditionally been used to measure the performance of production systems in terms of efficiency in converting inputs into outputs. In this paper, we present a novel use of the two-stage network DEA to evaluate the sustainable product design performances. While sustainable product design has been considered as one of the most important practices for achieving sustainability, one challenge faced by decision makers in both the private and public sectors is how to deal with the difficult technical trade-offs between traditional and environmental attributes which require new design concepts and engineering specifications. To deal with this challenge, we conceptualize "design efficiency" as a key measurement of design performance in terms of how well multiple product specifications and attributes are combined in a product design that leads to lower environmental impacts or better environmental performances. A two-stage network DEA model is developed for sustainable design performance evaluation with an "industrial design module" and a "bio design module." To demonstrate the applications of our DEA-based methodology, we use data of key engineering specifications, product attributes, and emissions performances in the vehicle emissions testing database published by the U.S. EPA to evaluate the sustainable design performances of different automobile manufacturers. Our test

C. Chen (✉)
Department of Business Administration, National Taiwan University, Taipei, Taiwan, R.O.C.
e-mail: cchen026@ntu.edu.tw

J. Zhu
International Center for Auditing and Evaluation, Nanjing Audit University, Nanjing, P.R., China

School of Business, Worcester Polytechnic Institute, Worcester, MA 01609, USA
e-mail: jzhu@wpi.edu

J.-Y. Yu
Department of Business Administration, National Taiwan University, Taipei, Taiwan, R.O.C.
e-mail: jyyu@ntu.edu.tw

H. Noori
Laurier School of Business, Wilfrid Laurier University, Waterloo, ON N2L 3C5, Canada
e-mail: hnoori@wlu.ca

J. Zhu (ed.), *Data Envelopment Analysis*, International Series in Operations Research & Management Science 238, DOI 10.1007/978-1-4899-7684-0_10

results show that sustainable design does not need to mean compromise between traditional and environmental attributes. Through addressing the interrelatedness of subsystems in product design, a firm can find the most efficient way to combine product specifications and attributes which leads to lower environmental impacts or better environmental performances. We also demonstrate how two-stage network DEA can be used to develop a analytical framework for evaluating sustainable design performances as well as to identify the most eco-efficient way to achieve better environmental performances through product design.

Keywords Design for the environment • Network DEA • Design performance evaluation

10.1 Introduction

Example 1: In December 2007, the European Commission proposed the introduction of legally-binding fuel efficiency standards for new cars. The proposed law says that CO_2 limits should be differentiated according to the type of car and that the so-called "utility parameter" used to define the targets should be the car's weight. In simple terms the proposal says heavier cars should get easier (higher) CO_2 standards and lighter cars should get tougher (lower) ones (European Federation for Transport and Environment 2008). Similarly, the U.S. Environmental Protection Agency (EPA) and the Department of Transportation's National Highway Traffic Safety Administration (NHTSA) are finalizing a set of fleet-wide average CO_2 emission standards where each vehicle has a different CO_2 emissions compliance target depending on its "footprint value" related to the size of the vehicle (U.S. EPA 2010).

Example 2: In the 2008 CTI Symposium on Automotive Transmissions, Robert Lee, Chrysler's vice president of power train engineering whose company's has not introduced any hybrid electric vehicles (HEVs), argued that, while a portion of the car-buying public may try to ride out that storm in HEVs, which now get most of the popular attention in the debate about how to minimize fuel consumption, engineers can still improve the conventional vehicles, too, by scouring them for fuel-wasting losses. "Solving the fuel economy puzzle requires a total-vehicle solution," Lee said, emphasizing the interrelatedness of the various vehicle subsystems (Design News 2008).

Sustainable product design or, sometimes equivalently, design for the environment (DfE), is considered as one of the most important practices for achieving sustainability. In recent years, however, there has been a fundamental shift in the ways which sustainable design performances are measured in both the public and private sectors, from emphasizing the absolute environmental performance to the eco-efficient design performance with carefully combined functional and environmental attributes (e.g., vehicle weight and fuel efficiency) in a product design (Ulrich and Eppinger 2012), as exemplified in the above two cases of applications. In this paper, we conceptualize the novel notion of "design efficiency" for combining multiple subsystems in the design process, and propose a comprehensive research framework based on the two-stage network Data Envelopment Analysis (DEA) for sustainable design performance evaluation. The vehicle emissions testing database published by the U.S. EPA (2009) will be used to demonstrate the applications of our proposed methodology in both the public sector (for evaluating the sustainable design performances of different vehicle designs by automakers) and private sector (for identifying the most eco-efficient sustainable design choices).

Today sustainable product design has received significant attention from both the public and private sectors worldwide. According to the most recent Green Brand Survey (2012) of 9000 consumers in Australia, Brazil, China, France, Germany, India, the U.S., and U.K., nearly three-quarters of consumers surveyed said that greenness was an important factor in determining which products to buy, and, 38 % of consumers rank “a socially responsible company” as a “very important” factor to consider when choosing products or services. In response to the strong public interest in sustainable purchasing, various directives and regulations aimed to encourage sustainable design practices have been considered or imposed by governments around the world. For example, in order to reduce greenhouse gas emissions and to achieve national energy independence, on July 29, 2011, U.S. President Barack Obama announced an agreement with thirteen large automakers to increase fuel economy to 54.5 miles per gallon for cars and light-duty trucks by model year 2025. In Europe, the Waste Electrical and Electronic Equipment Directive (WEEE) and Restriction of Hazardous Substances Directive (RoHS), which have gone into effect since 2006 in most EU member states, both require the “producer-polluter” to take the responsibility of processing and recycling electronic equipment when it reaches end-of-life to induce the producer to implement various practices for sustainable design (Lauridsen and Jørgensen 2010). In China, due to the increasing number of motor vehicles in recent years, the State Environmental Protection Administration has adopted the Euro IV standard, which is considered as a rather stringent emission standard for developing countries, since 2010 in order to induce Chinese automakers to make more significant efforts in designing and producing vehicles with low greenhouse gas emissions and carbon footprints (CNTV 2010).

Despite the calls and regulatory pressures from the general public and the governments, today’s companies have mixed responses regarding the implementation of sustainable design practices. On the one hand, most companies recognize the importance of sustainable design as exemplified by the fact that the websites of most Fortune 500 companies now feature an environmental section with substantial information regarding each company’s “commitment” to sustainable design. On the other hands, with only a handful of exceptions, most major companies still adopt a relatively reactive approach to sustainable product design. In the United States and Europe, the new CAFE Standard and the WEEE and RoHS Directives have all encountered rather strong resistance from the industries due to the potential technological and financial difficulties to achieve the required environmental performances. The fact is that, to design a product with improved environmental performance, a company usually needs to deal with some difficult technical trade-offs with new product specifications (Hopkins 2010). For example, a product made from 100 % recycled materials may have poor material consistency and durability (Malloy 1996; Verhoef et al. 2004). The zero-emission electric vehicles introduced in California in the early 1990s had rather poor traditional performances such as engine power, range, and size. These types of "green" products, which have been shown to have little chance to achieve market success, represent an inefficient use of resources in product design as the excellent environmental performances are combined with (or at the expense of) poor traditional product performances. According to the 2012 *Green Brand Survey*, mentioned previously, "offering good value" is still a predominant criterion for most consumers (75 % of respondents) in making a purchasing decision. As a result, the traditional performance measures for sustainable design which mostly focus on the "absolute scale" of environmental performance may not be sufficient to provide the industries with enough incentive to implement the practice of design for the environment as well as to offer consumers with adequate choices of well-functioning products with satisfactory levels of both traditional and environmental performances.

The purpose of the paper is to propose a methodology with the use of "design efficiency" as a novel measurement of sustainable design performances based on an innovative application of Data Envelopment Analysis, a method which has been widely applied to evaluate the efficiencies of decision-making units (DMUs). We conceptualize "design efficiency" as a key measurement of design performances, and develop a two-stage network DEA model for evaluating the sustainable design performances to find the most efficient way to combine product specifications and attributes to achieve better environmental performances through product design. We also discuss how to use the centralized and non-cooperative game theoretic models to solve for design efficiencies under the simultaneous, proactive, and reactive strategies adopted by firms for sustainable design. To demonstrate the applications of our proposed methodology, we use data of key engineering specifications, product attributes, and emissions performances in the vehicle emissions testing database published by the U.S. Environmental Protection Agency (EPA) to evaluate the sustainable design performances of different automobile manufacturers.

Our proposed performance measure of design efficiency, which measures how well multiple product specifications and attributes are combined in a product design to achieve better environmental performance, differs significantly from the traditional measures for sustainable design which mostly focus on the absolute scale of environmental performance. For example, if two products which are evenly matched in all the major functionalities (portability, material consistency, durability, etc.) generate different amounts of toxins, the product that generates a lower amount of toxins represents a more efficient design which leads to better environmental performance with the same input resources. Similarly, everything else being equal, if two motor vehicles with different sizes lead to the same level of greenhouse gas emissions, the vehicle with the larger size represents a more efficient design with a better combination of traditional and environmental product attributes. Notice that the use of design efficiency as an additional performance measure for sustainable design does not mean that one should ignore the traditional absolute measures of environmental performances which are often directly tied to the human or ecological impacts of a product. Rather, through better understanding of the design efficiency in sustainable design, a firm may utilize its limited resources in a more efficient way to design a product with better environmental performance, or, conversely, to meet the same environmental standard with a more efficient product design with a better combination of traditional and environmental attributes. Such an efficient sustainable design process would ultimately lead to the more efficient allocation of design resources for a firm as well as better product choices with improved functionalities and environmental performances for consumers.

The remainder of the paper is organized as follows. In Sect. 10.2, we review relevant literature. In Sect. 10.3, we conceptualize design processes to develop a research framework for measuring design efficiencies based on a two-stage network DEA model. In Sect. 10.4, we perform DEA analysis with the vehicle emissions testing database published by U.S. EPA to demonstrate how to use our proposed methodology for evaluating sustainable design performances. Test results are discussed in Sect. 10.5, and concluding remarks are in Sect. 10.6.

10.2 Literature Review

There exists a growing body of work on sustainable product design and DfE. One stream of the research focuses on micro-economic analyses of different sustainable design practices. Calcott and Walls (2000) compare the effects of different policy instruments that are used to target green design practices. Under a framework with design trade-offs, Chen (2001) analyzes the green product design decisions under different strategic and regulatory settings. With the introduction of recyclability through a technological parameter, Fullerton and Wu (1998) examine the effects of different public policies in a general equilibrium model. Atasu et al. (2008) model consumer's heterogeneous reaction to environmental friendliness through their differential appreciation for a remanufactured product versus a new product. Atasu and Souza (2010) investigate the impact of product recovery on design quality choices. While these papers provide good insights to a number of operational, strategic, and policy issues related to DfE, there has been less sustained work on performance measurement and evaluation for sustainable design.

Another stream of research in sustainable product design and DfE provides practical guidelines for implementing sustainable design practices. Handfield et al. (2001) propose a comprehensive conceptual framework with detailed implementation processes for DfE that connects corporate environmental objectives, design processes, and outcome evaluation. By using the framework of scenario planning, Noori and Chen (2003) propose a methodology for developing breakthrough products with environmental attributes. Fiksel (2009) uses case studies from major corporations to details implementation steps for DfE in the context of product life-cycle management. By using the industrial ecology principles and case studies, Graedel and Allenby (2009) identify a number of practical approaches to green design decisions. Ulrich and Eppinger (2012) discuss how to frame DfE as a material problem to provide incremental design solutions through the product "industrial" life cycle and the natural "bio" life cycle. The quantitative evaluation of sustainable product design performance, however, is an area which has not received much attention. Conway-Schempf and Lave (1999) and Hendrickson et al. (2006) develop an input-output approach for analyzing the life-cycle impact of a product which addresses several shortcomings of the traditional life-cycle assessment. Their approach, however, is primarily focused on the quantification of the environmental impacts of different products as opposed to using operations research techniques such as DEA to identify the efficient frontiers for evaluating and comparing different product designs as in our model.

Data Envelopment Analysis has been widely used to measure the performance of decision making units (DMUs) in terms of efficiency in combining inputs into outputs (Farrell 1957; Charnes et al. 1978; Liang et al. 2008b). Comprehensive reviews of research in DEA are provided in Emrouznejad et al. (2010) and Cook et al. (2010). Traditionally, DEA has been used in a single-stage model within a "black-box" framework. Such an efficiency measure, however, has limitations to deal with decision-making processes which can be divided into sub-processes or stages, where outputs of one sub-process are inputs to another sub-process. With a network structure, one might expect the decision maker to optimize the efficiencies of multiple sub-processes in a sequential fashion. To incorporate the multi-stage decision-making process into performance measurement, Färe and Grosskopf (1996, 2000) extend Shephard and Färe's (1979) production framework into a network DEA model. Recently, Kao and Hwang (2008) show that the whole-system efficiency can be decomposed into the product of sub-process efficiencies. Liang et al. (2008a) further develop two systematic approaches to analyze network efficiency: a game-theoretic non-cooperative approach and a centralized approach, which will be adopted in our model for evaluating the efficiency in sustainable product design.

In recent years, DEA has been increasingly used for performance evaluation in engineering design. Miyashita (2000) applies DEA to develop evaluation criteria to solve the collaborative design problem. Linton (2002) uses DEA to select materials which are efficient for various environmental indices. Farris et al. (2006) present a case study of how DEA is applied to generate objective cross-project comparisons for evaluating the relative performance of engineering design projects. By using DEA as a decision supporting tool, Cariaga et al. (2007) evaluate the degree to which each design alternative satisfies the customer requirements. Lin and Kremer (2010) apply DEA to solve the conceptual design problems and product family design problems. These papers, however, are all based on the standard DEA where the internal structure of a unit under evaluation is not modeled while our model is based on the more complex two-stage network DEA with a well-defined input-output internal structure.

In the literatures of management science and operations management, sustainable operations are commonly modeled as two-stage processes. Fleischmann et al. (1997) propose a framework decomposing the business logistics process into the "forward channel" and "reverse channel," which has been widely adopted in quantitative models of sustainable operations for product recovery and green supply chain management (Dekker et al. 2010). Recently, Ulrich and Eppinger (2012) propose a DfE framework with both the product "industrial" life cycle and natural "bio" life cycle. As in the above-mentioned analytical models for sustainable operations, our two-stage network DEA model allows decision makers to clearly identify the underlying factors and their interactions which lead to different environmental performances/consequences as well as the areas for future improvement. While the focus of the paper is not on DEA model building as our main analysis is largely based on Liang et al. (2008a), this paper proposes an innovative application of network DEA in sustainable product design with the following three major contributions. First, we conceptualize "design efficiency" as a key measurement of design performances, and develop a two-stage network DEA framework for evaluating the sustainable design performances to find the most efficient way to combine product specifications and attributes to achieve better environmental performances through product design. Second, we discuss how to use the centralized and non-cooperative game theoretic models to solve for design efficiencies under the simultaneous, proactive, and reactive strategies adopted by firms for sustainable design. Third, we demonstrate the innovative applications of our proposed methodology in both the private and public sectors by using data of key engineering specifications, product attributes, and emissions performances in the vehicle emissions testing database published by the U.S. Environmental Protection Agency (EPA) to evaluate the sustainable design performances of different automobile manufacturers. According to the authors' knowledge, our paper presents the first research work that applies network DEA to develop a comprehensive analytical framework with well-defined internal structure for analyzing the complex decision-making processes for sustainable design, which is of crucial importance to the future of human society, with empirical validation. In the section that follows, we will present our analytical framework based on a network DEA model.

10.3 Research Framework

We now present an analytical framework that integrate the structures of a two-stage network DEA model and a two-stage process commonly adopted in the existing literatures of sustainable operations and DfE. The proposed network DEA model for sustainable design performance evaluation includes two internal stages: an "industrial design process" and a "bio design process," as illustrated in Fig. 10.1, which are corresponding to the "Product Industry Life Cycle" and "Natural Bio Life Cycle" for sustainable design innovation presented in Hopkins (2010). Following Liang et al. (2008a), we consider a set of n different designs of a particular product as the decision making units (DMUs). Assume that for each product design, denoted by $DMU_j(j = 1, \ldots, n)$, there are m relevant engineering specifications as the inputs, denoted by $x_{ij}(i = 1, \ldots, m)$, to Stage 1 (industrial design) and D product attributes as outputs, denoted by $z_{dj}(d = 1, \ldots, D)$, from that stage.

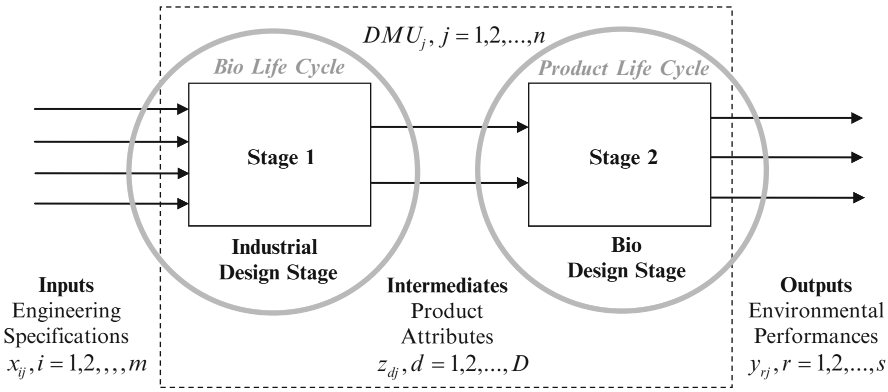


Fig. 10.1 The network DEA model for sustainable product design

These D outputs (attributes) are also the inputs to Stage 2 (bio design) and will be referred to as intermediate measures. The outputs from Stage 2, denoted by $y_{rj}(r = 1, 2, \ldots, s)$, are the levels of environmental performances of the product. We can then model the sustainable design problem with a two-stage network DEA model with either the "centralized" (integrated) approach in which efficiencies in both stages are optimized simultaneously or the "non-cooperative" (sequential) approach in which efficiencies in the two stages are optimized sequentially in any given order (Stage 1 first followed by Stage 2, or in the reverse order) (Liang, Cook, and Zhu 2008a). Notice that using the proposed two-stage model does not require that the industrial design and bio design processes be conducted separately. In fact, the centralized approach allows the simultaneous, joint decision-making for the industrial design and bio design processes concurrently. The major components of the proposed DEA framework are described below.

10.3.1 Stage 1: Industrial Design Performance

At the first stage, we evaluate the efficiency of the industrial design module, which can be viewed as the standard design process for combining engineering specifications (inputs) into product attributes (outputs). An engineering specification is defined as "a precise description of an engineering characteristic incorporated in a product design" (Ulrich and Eppinger 2012) and a product attribute is defined as "one of the main physical features of a product as the combination of a number of engineering specifications" (Noori 1990; Urban and Hauser 1993). For example, the portability of a handheld music device (product attribute) is determined by the combination of several engineering specifications, such as materials used, battery type, and the size of internal hard drive (or flash memory). The fuel economy of a vehicle (product attribute) is influenced by the combined effects of a number of engineering characteristics such as vehicle horsepower and engine compression ratio. The process of linking engineering characteristics with product attributes at the first stage is analogous to that used in standard methods for product design such as "House of Quality" and "Quality Function Deployment" (Hauser and Clausing 1988; Urban and Hauser 1993). The DEA analysis, however, allows us to evaluate the efficiency of resource usage of each product design (DMU) in terms of combining inputs (engineering specifications) into outputs (product attributes) in the industrial design process.

10.3.2 Stage 2: Bio design Performance

At Stage 2, we evaluate the efficiency of the bio design module by examining the links between key product attributes and environmental performances/ consequences. It is well documented that reducing the environmental impacts of a product through product design usually requires systematic design solutions to address the combined effects and interfaces of multiple product attributes (Hendrickson et al. 2006; Fiksel 2009). Based on the DfE concept proposed by Ulrich and Eppinger (2012), sustainable design and innovation is fundamentally a "material problem" which often requires redesign and reengineer of a product as well as its supply-chain functions to reduce the amount of toxins, the use of non-renewable resources, and the use of energy. Therefore, reducing the environmental impacts of a product involves not only the environmental attributes but also many of the traditional attributes. For example, if a company wants to reduce the use of virgin materials in a product, it is often necessary to redesign and reengineer the entire product so that it works properly (e.g., with the same material consistency and durability) and looks great without some of the virgin materials used in the original design (Hopkins 2010). Similarly, to reduce the emissions levels of a vehicle, a company usually needs to deal with the combined effects of a number of traditional and environmental attributes such as the size/ weight and the fuel economy. As another example, the exact amount of e-waste generated by a laptop computer is usually influenced by the combined effects of its recyclability and other product attributes such as size, weight, and portability. Therefore, the DEA analysis at the second stage is aimed to evaluate the efficiency of a product design (as a DMU) in combining key attributes to reduce the environmental impacts or to improve the environmental performances of a product.

Depending on data availability, the outputs from the second stage can be either the life-cycle environmental impacts (Hendrickson et al. 2006; Fiksel 2009) or only one or a few environmental performances or impacts of interest, such as the amounts of e-waste and levels of vehicle emissions. We keep our model general with the understanding that, while reducing the overall life-cycle environmental impacts should be the ultimate goal of sustainable design, the DfE approach with incremental improvements on one or a few environmental performances is usually more executable for most businesses today (Hopkins 2010). Analyses based on only one or a few environmental performances or impacts of interest can also be parts of or integrated into the more complete environmental assessment under the life-cycle framework.

10.3.3 Design Performance Evaluation

We now discuss the performance measures for each of the two stages (industrial design and bio design) as well as the overall two-stage network model. On the basis of Charnes et al. (1978), the efficiencies of the first and second stages for a DMU_j $(j = 1,\ 2,\ \ldots, n)$ can be calculated as:

$$e_j^1 = \frac{\sum_{d=1}^{D} w_d z_{dj}}{\sum_{i=1}^{m} v_i x_{ij}} \quad \text{and} \quad e_j^2 = \frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{d=1}^{D} \widetilde{w}_d z_{dj}} \tag{10.1}$$

where v_i, w_d, $\widetilde{w}_d$, and u_r are unknown non-negative weights to be solved. These ratios are then used in a mathematical programming problem which can be converted into a linear program. It is noted that w_d is set equal to $\widetilde{w}_d$ as in Liang et al. (2008a).

Two different approaches, termed "centralized" approach and "non-cooperative" (decentralized) approach, can be used to measure the efficiencies of each of the two individual stages as well as the overall two-stage process (Liang et al. 2008a). With the centralized approach, the efficiencies of both stages (industrial design and bio design) are evaluated simultaneously to determine a set of optimal weights on the intermediate measures that maximizes the aggregate or global efficiency score in a joint decision-making process. With the decentralized approach based on the leader-follower paradigm of the Stackelberg model, Stage 1 (industrial design) is the leader whose performance (efficiency) is more important and thus optimized first. Then the efficiency of Stage 2 (bio design) as the follower is computed, subject to the requirement that the leader's efficiency remains fixed. Similarly, with the decentralized approach based on the follower-leader paradigm, Stage 2 (bio design) is the leader and optimized first while Stage 1 (industrial design) is the follower. The mathematical program for solving the efficiencies under the centralized approach used in our DEA test is described as follows: For a specific DUM_o, the following linear programming model maximizes the centralized (overall) efficiency as the product of individual efficiencies of the two stages.

$$Max \;\; e_o^1 \bullet e_o^2 = \frac{\sum_{r=1}^{s} u_r y_{ro}}{\sum_{i=1}^{m} v_i x_{io}} \tag{10.2}$$

$$s.t. \;\; e_j^1 \le 1 \quad \text{and} \quad e_j^2 \le 1 \quad \text{and} \quad w_d = \widetilde{w}_d$$

Model (10.2) can be converted into the following linear program

$$
\begin{aligned}
& Max \quad \sum_{r=1}^{s} u_r y_{ro} \\
& s.t. \ldots \quad \sum_{r=1}^{s} u_r y_{rj} - \sum_{d=1}^{D} w_d z_{dj} \le 0 j = 1, 2, \ldots, n \\
& \sum_{d=1}^{D} w_d z_{dj} - \sum_{i=1}^{m} v_i x_{ij} \le 0 j = 1, 2, \ldots, n \\
& \sum_{i=1}^{m} v_i x_{io} = 1 \\
& w_d \ge 0, \; d = 1, 2, \ldots, D; \; v_i \ge 0, \; i = 1, 2, \ldots, m; \; u_r \ge 0, \; r = 1, 2, \ldots, s
\end{aligned}
\tag{10.3}
$$

We can then obtain the efficiencies for the first and second stages, namely

$$e_o^1 = \frac{\sum_{d=1}^{D} w_d^* z_{do}}{\sum_{i=1}^{m} v_i^* x_{io}} = \sum_{d=1}^{D} w_d^* z_{do} \quad \text{and} \quad e_o^2 = \frac{\sum_{r=1}^{s} u_r^* y_{ro}}{\sum_{d=1}^{D} w_d^* z_{do}}. \tag{10.4}$$

Similarly, the mathematical programs used under the non-cooperative approach can be established. To obtain the overall efficiency of the two-stage process, let e_o^{1*} and e_o^{2*} denote the efficiencies of the first and second stages obtained with the centralized or non-cooperative approach. The overall two-stage efficiency, denoted by $e_{\mathrm{o}}^{\mathrm{all}*}$, can then be calculated as the product of the individual efficiencies of the two stages (i.e., $e_{\mathrm{o}}^{\mathrm{all}*} = e_o^{1*} \times e_o^{2*}$) regardless of whether the centralized or non-cooperative approach is used, as shown in Liang et al. (2008a).

With the solved individual and overall efficiencies from the DEA model, we will be able to compare and evaluate the design performances of different DMUs (product designs), and identify the most efficient way to combine product specifications and attributes to achieve better environmental performances or to reduce environmental impacts through product design. Note that, in a typical DEA model for measuring production efficiency, an efficient DMU is the one that is capable of using the minimum input resources to produce the same levels of outputs or, equivalently, using the same input resources to produce the maximum levels of outputs. Similarly, in our DEA model for measuring sustainable design performances, an efficient DMU (product design) is the one that has the best combination of engineering specifications or product attributes to achieve the same environmental performances (i.e., how well different product specifications and attributes are combined in a product design to achieve the environmental performances) or, equivalently, the DMU (product design) that is capable of using the same levels of product specifications/attributes to achieve the best environmental performances. Notice that, while the models in Kao and Hwang (2008) and Liang et al. (2008a) are developed under the assumption of geometric mean of two stages' efficiency scores, additive forms of efficiency decomposition can also be used. For example, Chiou et al. (2010) develop an integrated DEA model where the overall efficiency is defined as a (weighted) average efficiency of two stages under the assumptions of constant and variable returns to scale (CRS and VRS). Chen et al. (2009) discussed the additive efficiency decomposition under both CRS and VRS assumptions when a set of DMU-related weights are used. Our proposed framework can readily be applied to the models proposed in the above studies.

10.3.4 Strategic Implications

The two-stage network DEA model and different solution approaches presented above make it possible to analyze three different sustainable design strategies firms may adopt, namely:

1. Simultaneous Approach: A firm simultaneously optimizes both the industrial and bio design processes, which can be analyzed with the centralized approach of two-stage DEA.
2. Reactive Approach: A firm optimizes the industrial design process first, and then optimizes the bio design process, which can be analyzed with the decentralized approach of two-stage DEA with Stage 1 (industrial design) as the leader and Stage 2 (bio design) as the follower.
3. Proactive Approach: A firm optimizes the bio design process first, and then optimizes the industrial design process, which can be analyzed with the decentralized approach with Stage 2 (bio design) as the leader and Stage 1 (industrial design) as the follower.

With the proposed two-stage network DEA model, decision makers would be able to investigate and compare the individual and overall design performances with either the centralized and decentralized approach under different strategies for sustainable product design. It should be noted that, for a product with a simpler, single-stage design process, our analytical framework can be easily modified and reduced to a single-stage DEA model. In the section that follows, we will demonstrate the applications of our analytical model in evaluating the sustainable design performances with vehicle emissions testing data for the automobile industry.

10.4 Data Collection and Research Procedure

In this section, we use the data of product specifications, attributes, and indices of vehicle emissions performances in the vehicle emissions testing database published by the U.S. EPA (2009) to demonstrate the applications of our model in evaluating sustainable design performances of vehicles introduced in North American in 2009. Since the database only includes data of vehicle specifications, attributes, and emission performance indices which are considered relevant to the emissions tests by the agency as opposed to the complete data sets of vehicle specifications, attributes, and life-cycle environmental performances, the purpose of our analysis is to show how to use our model for evaluating sustainable design performances instead of suggesting the more eco-efficient product designs or assessing the actual design performances of automobile manufacturers. To present our analysis from the problem-solving perspective, we first identify two applications of our DEA tests in the public and private sectors corresponding to the two real-world examples regarding the new performance measures considered by the European Commission and U.S. EPA as well as different design options considered by Chrysler.

Application #1 (Public Sector): *An environmental protection agency has been using the tailpipe emissions levels as the primary measures of environmental performances of different automobile manufacturers for years. However, it has come to the agency's attention in recent years that many consumers do not purchase environmentally-friendly vehicles due to the perception that good environmental performances are usually at the expense of other important vehicle performances such as power and size. As a result, the agency would like to understand the overall "design efficiencies" of different automobile manufacturers in terms of combining engineering specifications and customer attributions in product design to achieve environmental performances as an alternative measure of environmental excellence.*

Application #2 (Private Sector): *A major automobile manufacturer is considering investing in developing hybrid electric engine in order to reduce emissions and to improve the environmental performances of its vehicles. However, some of the company's designers and engineers argue that good environmental performances can be achieved with the traditional ICE engine in an eco-efficiency way with carefully combined vehicle subsystems. Therefore, the company would like to study and compare its own vehicles as well as other ICE and hybrid vehicles offered by its competitors in terms of the "design efficiency" for achieving environmental performances.*

We now demonstrate how to use the proposed two-stage network DEA analysis to solve the problems for the environmental protection agency and private company.

10.4.1 Data

All new cars and trucks sold in the U.S. must be certified to meet federal emissions standards. This is accomplished by performing laboratory tests on pre-production vehicles by the U.S. EPA or by manufacturers at their own facilities under EPA's supervision. The 2009 database includes the data of 2885 new vehicles tested in the year. For each tested vehicle, the database provides information about the relevant engineering specifications, attributes, and vehicle emissions performances based on separate tests performed on city and highway. Since our purpose is to demonstrate the applications of the proposed model, we only analyze the emissions data based on the city tests. In many cases, multiple vehicles of the same model/option are tested. To analyze the efficiencies of individual vehicle designs, we first sort all the data by each "carline" identified by EPA as one major option of a particular vehicle model. For example, the two-wheel drive (2WD) option and four-wheel drive option (4WD) of Chrysler Grand Cherokee are considered as two different carlines. Similarly, the option with 5-speed manual transmission and the option with automatic transmission of Honda Civic are considered as two different carlines. For some carlines with repetitive data in the database, the average values of engineering specifications, attributes, and emissions levels are calculated whenever applicable. Missing data, however, are quite common in the database. We therefore remove some of the engineering specifications, attributes, and emissions test results with significant portions of missing data. Carlines with missing data are also removed from our analysis, which results in 534 carlines (product designs) used in our analysis with data of cubic inch displacement (CID), rated horsepower (RHP), compression ratio (cmp), axle ratio (axle), equivalent test weight (ETW), fuel economy (MPG), hydrocarbon emissions (HC), carbon monoxide emissions (CO), carbon dioxide emissions (CO_2), and nitrogen oxide emissions (NOx). These carlines are introduced by more than 20 different manufacturers including all the major automakers for the North American market such as Chrysler, Ford, General Motors, BMW, Mitsubishi, Mercedes Benz, Honda/Acura, Hyundai, Kia, Nissan/Infiniti, SAAB, Mazda, Toyota/Lexus, Audi/Volkswagen, and Volvo.

10.4.2 Testing Procedure

Our testing procedure follows Liang et al. (2008a) for two-stage network DEA with efficiency decomposition. Since we do not have the private information about whether the automobile manufacturers use the simultaneous, proactive or reactive strategy for sustainable design, we only use the centralized approach in our analysis to simultaneously evaluate the efficiencies of both stages to obtain the maximum overall efficiency, which is considered as the more "neutral" measurements of design performances. Our analysis can be easily modified with the decentralized approach if the exact information about the sustainable design strategy (proactive or reactive strategy) used by each manufacturer is available.

With the compiled 2009 vehicle emissions testing database, we consider that each carline forms a DMU as a particular product design with four relevant engineering specifications, namely, cubic inch displacement, rated horsepower, compression ratio, and axle ratio, as the inputs. Two attributes, fuel economy and equivalent test weight, which is commonly used as a surrogate measure of vehicle size (Crandall and Graham 1989; Chen and Zhang 2009), are considered as the intermediate measures. The levels of hydrocarbon emissions, carbon monoxide emissions, carbon dioxide emissions, and nitrogen oxide emissions, are considered as the outputs. In DEA, higher levels of outputs usually indicate better performance. Therefore, we treat the outputs by taking the reciprocals of the emission levels. Similarly, we use the reciprocals of cubic inch displacement, rated horsepower, compression ratio, and equivalent test weights in our analysis to fit the DEA use. Notice that the definitions of inputs, intermediates, and outputs in our network DEA analysis are similar to those of engineering characteristics, customer attributes, and product performances for the design process of car doors discussed in the "House of Quality" framework proposed by Hauser and Clausing (1988). In the first stage (Industrial Design Module), the interactions between engineering specifications (cubic inch displacement, rated horsepower, compression ratio, axle ratio) and customer attributes (size/weight and fuel economy) are analyzed. In the second stage (Bio Design Module), the effects of customer attributes on environmental impacts/performances (the emissions of hydrocarbon, carbon monoxide, carbon dioxide, and nitrogen oxide) are analyzed.

It should be noted that one needs to exert caution when dealing with both the ratio and raw data to avoid the situation where they are not properly mixed, but the use of ratio data (fuel economy) in our model does not lead to any of those problematic situations described in Dyson et al. (2001) and Cooper et al. (2007) (e.g., a factor appears on both the input and output sides). Also notice that the list of inputs does not include any human resources due to data availability. Human contribution to product design, such as knowledge and creativity, is usually hard to measure and quantify. While the number of patents is sometimes used as a surrogate to measure human contribution in the existing literature, such information (the number of patents used in each of the individual vehicle design) is not available in the database or in any other data sources. Another technical issue regarding the centralized approach for solving a two-stage network DEA model is that, while the optimal overall efficiencies for DMUs are unique, the individual efficiencies of the two stages may not be unique. Therefore, we use the procedure proposed in Liang et al. (2008a) to check for uniqueness of solved individual efficiencies, which shows that all the efficiency decompositions in our DEA analysis are unique.

10.5 Research Results

By using the data of 534 different carlines (DMUs) introduced in North America in 2009, we perform the DEA test with the procedure discussed in the previous section to obtain the design efficiencies of the two stages as well as the overall (centralized) efficiency. The results of the overall and individual design performances are presented and discussed below.

10.5.1 Overall Performance Comparison

Figure 10.2 presents the first-stage and second-stage efficiencies of all the carlines tested in our analysis. According to the figure, the first-stage efficiencies of most carlines are higher than the second-stage efficiencies; i.e., the "positions" of most carlines lie below the diagonal line of the diagram. In particular, the first-stage efficiencies of most carlines are higher than 50 % (0.5), while the second-stage efficiencies are mostly lower than 50 %. This indicates that, while most manufacturers are quite capable of combining engineering specifications to achieve satisfactory levels of vehicle weight (size) and fuel economy with relatively high design efficiencies at the first stage (industrial design), many of them are less capable of utilizing the resulting combinations of vehicle weight and fuel economy to produce

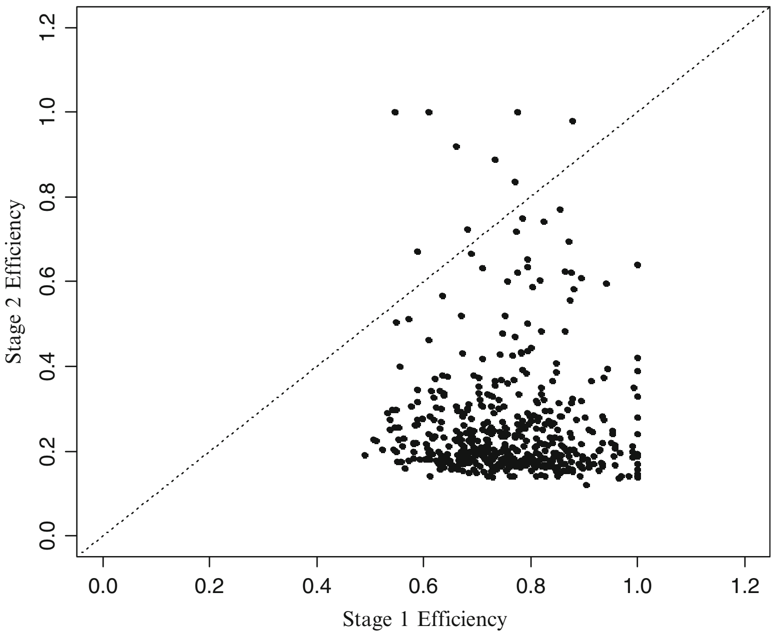


Fig. 10.2 A Comparison of Design Efficiencies of the Two Stages (All Carlines)

Table 10.1 A comparison of design efficiencies of selected manufacturers

Stage 1 Efficiency		Stage 2 Efficiency	Overall Efficiency
American Company 1	0.7066	0.4269	0.3033
American Company 2	0.7322	0.1990	0.1454
Japanese Company 1	0.7560	0.3048	0.2304
Japanese Company 2	0.7470	0.3577	0.2637
European Company 1	0.7716	0.2375	0.1831
European Company 2	0.7964	0.3752	0.2966

good emissions performances with relatively low design efficiencies at the second stage (bio design).

Due to space limitation, we will only present the detailed test data and results of six major companies with disguised names as American Company 1 (AC1), American Company 2 (AC2), Japanese Company 1 (JC1), Japanese Company 2 (JC2), European Company 1 (EC1), and European Company 2 (EC2). Table 10.1 shows the summary of the average first-stage efficiencies, second-stage efficiencies, and overall (centralized) efficiencies of the six manufacturers. According to the table, AC1 has the highest average overall efficiency. While the average first-stage efficiency (for industrial design) of AC1 is slightly lower than all the other manufacturers, its significantly higher second-stage efficiency (for bio design) not only offsets the relatively lower first-stage efficiency but also leads to the highest average overall efficiency among the six manufacturers. EC2, which has the highest first-stage efficiency but lower second-stage efficiency than AC1, ranks second for the overall efficiency. The two Japanese companies (JC1 and JC2), both with moderate first-stage and second-stage efficiencies, rank third and fourth for the overall efficiency. EC1 and AC2 rank fifth and sixth for the overall efficiency largely because of their significantly lower second-stage efficiencies.

10.5.2 Individual Carline Performance Comparison

We now present more detailed test data and results for the six automobile manufacturers. Due to space limitation, we will only present the test data and results of 23 selected carlines for each company, including the top three carlines with the highest overall efficiencies as well as 20 other commonly seen carlines as the representative examples. (We only present partial results since one company has more than 100 carlines, and three others have more than 45 carlines listed in the database.) Tables 10.2, 10.3, 10.4, 10.5, 10.6, and 10.7 show the overall, first-stage, and second-stage efficiencies as well as the data of inputs, intermediate measures, and outputs of the 23 selected carlines of AC1, AC2, JC1, JC2, EC1, and EC2, respectively. We first note that, carline #2 produced by JC1 (Table 10.4), a compact car which is one of the first hybrid vehicles introduced in North America, justifies its reputation as an environmentally friendly all-around vehicle with the highest

Table 10.2 Data and test results for selected carlines of American Company 1 (AC1)

Carline (DMU)	Overall Efficiency	Stage 1 Efficiency	Stage 2 Efficiency	CID	RHP	CMP	Axle	ETW	MPG	HC	CO	CO_2	NOx
1	0.18784	0.81114	0.23158	268.31	247.23	9.84	3.31	3923	19.32	0.0216	0.4108	329.77	0.0054
2	0.45475	0.75539	0.60200	140.00	160.00	9.70	4.04	3583	25.83	0.0067	0.1667	343.67	0.0033
3	0.17072	0.80975	0.21083	182.00	217.00	10.00	3.46	3750	22.30	0.0240	0.1800	397.00	0.0100
4	0.64428	0.77061	0.83606	182.00	215.50	10.00	3.46	3938	20.85	0.0075	0.0425	107.50	0.0050
5	0.14839	0.66017	0.22477	281.00	250.00	9.40	3.27	4500	15.50	0.0400	0.5700	485.50	0.0050
6	0.58829	0.78343	0.75091	161.00	205.50	10.00	3.82	3688	25.55	0.0055	0.1450	157.00	0.0050
7	0.56000	0.94118	0.59500	140.00	155.00	9.70	2.00	4250	37.40	0.0080	0.2300	237.00	0.0200
8	0.64000	1.00000	0.64000	140.00	155.00	9.70	2.00	4000	45.20	0.0070	0.1900	196.00	0.0100
9	0.20329	0.58709	0.34627	244.00	210.00	9.70	3.64	5250	16.10	0.0280	0.5300	283.00	0.0050
10	0.14161	0.61661	0.22967	280.00	292.00	9.80	3.55	5250	16.75	0.0260	0.3600	531.50	0.0150
11	0.12445	0.62840	0.19805	330.00	310.00	9.80	3.31	6000	15.27	0.0400	0.7467	522.33	0.0100
12	0.60996	0.60996	1.00000	280.00	262.67	9.40	3.69	5556	16.90	0.0062	0.1333	117.00	0.0011
13	0.15219	0.66414	0.22915	280.00	292.00	9.80	3.55	5000	17.90	0.0260	0.2100	495.00	0.0100
14	0.12534	0.67419	0.18591	330.00	310.00	9.80	3.31	6000	17.10	0.0390	0.9450	517.50	0.0100
15	0.26126	0.73360	0.35613	330.00	310.00	9.80	3.15	5375	17.85	0.0215	0.4450	250.50	0.0050
16	0.60652	0.65939	0.91983	330.00	310.00	9.80	3.53	5900	16.92	0.0132	0.2280	105.00	0.0020
17	0.13448	0.53680	0.25052	330.00	310.00	9.80	3.31	6000	11.60	0.0420	0.3500	532.00	0.0100
18	0.28129	0.60790	0.46272	330.00	310.00	9.80	3.15	5375	12.90	0.0145	0.2350	244.50	0.0050
19	0.54527	0.54527	1.00000	330.00	310.00	9.80	3.53	5900	12.22	0.0070	0.0840	102.80	0.0040
20	0.27440	0.74345	0.36909	140.00	171.00	9.70	4.13	3625	28.30	0.0110	0.2900	314.00	0.0100
21	0.13385	0.69798	0.19177	140.00	171.00	9.70	4.09	3875	23.75	0.0235	0.3350	374.00	0.0100
22	0.14476	0.75513	0.19170	140.00	143.00	9.70	4.01	3594	24.28	0.0225	0.2325	366.75	0.0100
23	0.19589	0.66249	0.29569	244.00	207.00	9.70	3.85	4250	17.60	0.0320	0.5767	334.00	0.0033

Notes: CID = cubic inch displacement, RHP = rated horsepower, cmp = compression ratio, axle = axle ratio, ETW = equivalent test weight (lb), MPG = fuel economy, HC = hydrocarbon emissions, CO = carbon monoxide emissions, CO_2 = carbon dioxide emissions, NOx = nitrogen oxide emissions

Table 10.3 Data and test results for selected carlines of American Company 2 (AC2)

Carline (DMU)	Overall Efficiency	Stage 1 Efficiency	Stage 2 Efficiency	CID	RHP	cmp	Axle	ETW	mpg	HC	CO	CO_2	NOx
1	0.17891	0.87979	0.20335	145.00	175.00	10.40	2.89	3750	27.80	0.0320	0.8700	323.00	0.0100
2	0.15866	0.93311	0.17003	204.00	185.00	9.50	2.48	4000	21.30	0.0365	0.6850	417.00	0.0100
3	0.61158	0.82505	0.74127	231.00	200.00	9.40	3.05	3875	21.80	0.0060	0.2200	406.00	0.0100
4	0.13488	0.71150	0.18957	145.00	163.00	10.40	3.91	4000	23.90	0.0430	1.4900	371.00	0.0100
5	0.13858	0.84916	0.16320	122.00	260.00	9.20	3.73	3375	23.30	0.0410	0.2600	381.00	0.0400
6	0.13728	0.84596	0.16228	122.00	260.00	9.20	3.73	3375	24.30	0.0420	0.4600	367.00	0.0100
7	0.17398	0.91749	0.18962	134.00	148.00	10.00	3.77	3042	31.60	0.0350	1.0467	280.00	0.0100
8	0.13117	0.78808	0.16645	279.00	275.00	10.00	3.11	4250	19.00	0.0420	0.4800	467.00	0.0100
9	0.20414	0.78188	0.26109	237.00	236.00	9.80	2.93	4250	19.63	0.0293	0.8433	405.33	0.0033
10	0.17483	0.86624	0.20182	110.00	138.00	10.50	4.03	3125	31.00	0.0180	0.2950	287.00	0.0150
11	0.19163	0.75497	0.25382	145.00	170.00	10.40	3.63	3875	33.20	0.0450	0.9100	266.00	0.0200
12	0.15308	0.87600	0.17475	218.00	250.00	10.20	2.77	4000	21.40	0.0490	1.1500	404.00	0.0100
13	0.54485	0.87595	0.62201	214.00	211.00	9.80	2.86	3875	22.90	0.0070	0.2500	387.00	0.0100
14	0.16421	0.71933	0.22828	218.00	288.00	11.30	3.16	5000	20.60	0.0240	0.5300	454.00	0.0200
15	0.14055	1.00000	0.14055	376.00	620.00	8.80	3.42	3625	16.80	0.0880	2.6800	532.00	0.0300
16	0.22273	0.72576	0.30689	122.00	210.00	9.50	3.55	4000	23.60	0.0150	0.2000	382.00	0.0200
17	0.11676	0.70534	0.16554	364.00	395.00	10.90	4.10	5000	14.90	0.0460	0.8800	617.00	0.0100
18	0.14756	0.63717	0.23159	325.00	300.00	9.90	4.10	5500	15.80	0.0445	0.9050	559.50	0.0050
19	0.12624	0.55275	0.22839	325.00	320.00	9.90	3.42	6000	12.60	0.0330	0.5600	502.00	0.0100
20	0.10447	0.69831	0.14960	380.00	400.00	10.50	3.42	5500	15.40	0.0930	1.5800	597.00	0.0200
21	0.13421	0.74391	0.18041	325.00	300.00	9.90	3.42	5000	18.40	0.0560	1.5500	464.00	0.0100
22	0.12750	0.69084	0.18456	178.00	190.00	10.00	3.73	4250	21.70	0.0390	1.3100	406.00	0.0100
23	0.16038	0.68752	0.23327	325.00	320.00	9.90	3.08	5250	15.78	0.0280	0.5400	512.50	0.0050

Notes: CID = cubic inch displacement, RHP = rated horsepower, cmp = compression ratio, axle = axle ratio, ETW = equivalent test weight (lb), MPG = fuel economy, HC = hydrocarbon emissions, CO = carbon monoxide emissions, CO_2 = carbon dioxide emissions, NOx = nitrogen oxide emissions

Table 10.4 Data and test results for selected carlines of Japanese Company 1 (JC1)

Carline (DMU)	Overall Efficiency	Stage 1 Efficiency	Stage 2 Efficiency	CID	RHP	cmp	Axle	ETW	MPG	HC	CO	CO_2	NOx
1	0.34800	0.99321	0.35038	91.00	106.00	10.50	3.98	2625	37.40	0.0128	0.0750	157.83	0.0100
2	0.86034	0.87879	0.97900	91.00	76.00	13.00	4.11	3250	66.60	0.0080	0.0200	131.00	0.0100
3	0.17962	0.86263	0.20822	110.00	132.00	10.00	4.22	3000	34.50	0.0280	0.1950	256.50	0.0200
4	0.16894	0.80805	0.20907	110.00	132.00	10.00	4.28	3188	32.93	0.0303	0.2125	268.00	0.0300
5	0.17780	0.87419	0.20339	110.00	128.00	10.00	4.04	3000	34.10	0.0310	0.3050	257.00	0.0250
6	0.20316	0.78645	0.25833	144.00	157.25	9.80	3.53	3688	26.95	0.0173	0.1525	325.50	0.0200
7	0.16126	0.89756	0.17966	144.00	161.00	9.80	3.49	3250	25.80	0.0245	0.2150	343.00	0.0150
8	0.22169	0.81315	0.27263	211.00	268.00	10.80	3.69	3875	24.50	0.0190	0.1200	360.00	0.0100
9	0.20006	0.78837	0.25376	211.00	272.00	10.80	3.69	4000	23.80	0.0190	0.1700	366.00	0.0100
10	0.30855	0.78489	0.39311	152.00	204.00	12.00	4.10	4000	24.60	0.0225	0.0650	357.00	0.0150
11	0.55519	0.77164	0.71949	211.00	292.00	11.80	3.77	4500	27.90	0.0070	0.0900	313.00	0.0100
12	0.22646	0.83452	0.27136	211.00	303.00	11.80	3.77	4000	23.85	0.0290	0.1200	370.50	0.0200
13	0.15585	0.74966	0.20789	262.00	288.00	10.50	3.77	4250	20.30	0.0240	0.2750	432.00	0.0150
14	0.20747	0.82257	0.25223	281.00	380.00	11.80	2.94	4750	20.20	0.0220	0.1850	433.50	0.0100
15	0.35105	0.93653	0.37484	303.00	416.00	11.80	2.94	4250	19.65	0.0135	0.0950	444.50	0.0100
16	0.22758	0.90872	0.25044	152.00	180.00	10.40	2.74	3750	27.30	0.0190	0.1500	325.00	0.0100
17	0.26889	0.73249	0.36709	211.00	270.00	10.80	3.29	4500	21.10	0.0220	0.0800	414.00	0.0200
18	0.16334	0.71622	0.22806	211.00	270.00	10.80	3.48	4500	22.10	0.0280	0.2100	399.00	0.0200
19	0.17663	0.75144	0.23506	211.00	266.00	10.80	3.08	4500	22.40	0.0350	0.2200	395.00	0.0100
20	0.12941	0.67872	0.19067	241.00	236.00	10.00	3.73	4500	20.40	0.0320	0.3700	432.00	0.0100
21	0.27719	0.54829	0.50555	285.00	276.00	10.00	4.11	6000	17.15	0.0140	0.1000	255.50	0.0100
22	0.20164	0.61910	0.32570	346.00	383.00	10.20	3.91	6500	15.30	0.0200	0.1800	575.00	0.0150
23	0.29616	0.70857	0.41797	346.00	381.00	10.20	4.30	5563	17.35	0.0133	0.1425	380.75	0.0125

Notes: CID = cubic inch displacement, RHP = rated horsepower, cmp = compression ratio, axle = axle ratio, ETW = equivalent test weight (lb), MPG = fuel economy, HC = hydrocarbon emissions, CO = carbon monoxide emissions, CO_2 = carbon dioxide emissions, NOx = nitrogen oxide emissions

Table 10.5 Data and test results for selected carlines of Japanese Company 2 (JC2)

Carline (DMU)	Overall Efficiency	Stage 1 Efficiency	Stage 2 Efficiency	CID	RHP	cmp	Axle	ETW	MPG	HC	CO	CO_2	NOx
1	0.17512	0.87118	0.20101	91.00	117.00	10.40	4.59	2844	35.75	0.0168	0.2450	248.00	0.0100
2	0.35606	0.80159	0.44419	82.00	110.00	10.80	4.94	3125	54.60	0.0080	0.1300	163.00	0.0100
3	0.28454	0.87816	0.32401	114.00	154.50	11.00	4.52	3146	28.77	0.0152	0.2350	169.17	0.0067
4	0.27213	0.75563	0.36014	144.00	185.67	10.50	4.42	3708	26.84	0.0126	0.0944	329.00	0.0189
5	0.21965	0.77588	0.28310	144.00	201.00	11.00	4.60	3750	26.00	0.0175	0.1050	342.50	0.0200
6	0.22902	0.80633	0.28403	212.00	280.00	11.20	4.31	4000	22.60	0.0170	0.1400	393.00	0.0200
7	0.23029	0.79289	0.29044	212.00	271.00	10.00	3.55	3875	20.70	0.0170	0.1000	427.00	0.0200
8	0.51893	0.79411	0.65346	212.00	271.00	10.50	4.31	3938	23.83	0.0068	0.1200	373.50	0.0150
9	0.50382	0.79263	0.63563	212.00	271.00	10.50	4.31	3875	23.70	0.0070	0.1000	373.50	0.0150
10	0.28322	0.76849	0.36854	224.00	305.00	11.20	4.53	4250	21.20	0.0140	0.1100	421.00	0.0100
11	0.23926	0.72280	0.33102	224.00	300.00	11.20	4.53	4500	20.00	0.0190	0.0900	443.00	0.0200
12	0.14689	0.60916	0.24114	140.00	240.00	8.80	4.53	4250	20.90	0.0200	0.3700	426.00	0.0100
13	0.22639	0.70396	0.32159	144.00	166.00	9.70	4.50	3750	25.30	0.0140	0.1200	349.00	0.0200
14	0.20331	0.68125	0.29844	144.00	166.00	9.70	4.50	3875	24.75	0.0155	0.1400	359.00	0.0200
15	0.49248	0.68058	0.72361	144.00	166.00	9.70	4.63	3875	24.00	0.0060	0.3350	369.50	0.0150
16	0.34763	0.66964	0.51913	144.00	166.00	9.70	4.63	3938	23.50	0.0085	0.4100	376.50	0.0150
17	0.23042	0.61961	0.37188	212.00	250.00	10.00	4.53	4750	18.60	0.0160	0.1000	476.00	0.0100
18	0.18837	0.62248	0.30261	212.00	242.50	10.25	4.31	4875	20.45	0.0195	0.1450	435.00	0.0150
19	0.20220	0.65991	0.30640	212.00	250.00	10.50	4.31	4625	20.75	0.0240	0.1100	428.00	0.0250
20	0.24228	0.64456	0.37589	212.00	250.00	10.50	4.31	4750	20.00	0.0180	0.0800	443.00	0.0200
21	0.19412	0.67336	0.28828	224.00	300.00	11.00	4.53	4750	18.90	0.0240	0.1200	469.00	0.0200
22	0.14684	0.89205	0.16461	122.00	197.00	11.00	4.76	3250	25.60	0.0490	0.5900	344.00	0.0300
23	0.13190	0.96410	0.13681	132.00	237.00	11.10	4.95	3125	22.20	0.0470	0.7700	398.00	0.0200

Notes: CID = cubic inch displacement, RHP = rated horsepower, cmp = compression ratio, axle = axle ratio, ETW = equivalent test weight (lb), MPG = fuel economy, HC = hydrocarbon emissions, CO = carbon monoxide emissions, CO_2 = carbon dioxide emissions, NOx = nitrogen oxide emissions

Table 10.6 Data and test results for selected carlines of European Company 1 (EC1)

Carline (DMU)	Overall Efficiency	Stage 1 Efficiency	Stage 2 Efficiency	CID	RHP	cmp	Axle	ETW	MPG	HC	CO	CO_2	NOx
1	0.16177	0.69116	0.23406	121.00	200.00	9.60	4.08	3958	22.67	0.0195	0.2617	392.67	0.0250
2	0.21183	0.73130	0.28966	121.00	200.00	9.60	4.06	3750	27.45	0.0145	0.4100	323.50	0.0150
3	0.17770	0.82271	0.21600	121.00	200.00	9.60	3.94	3375	26.05	0.0175	0.5750	340.50	0.0100
4	0.34554	0.79332	0.43556	121.00	200.00	9.60	3.94	3500	26.00	0.0090	0.5050	341.00	0.0150
5	0.32735	0.76597	0.42737	121.00	200.00	9.60	3.94	3625	25.85	0.0095	0.5600	343.00	0.0150
6	0.15570	0.84309	0.18468	121.00	200.00	9.60	3.09	3625	26.00	0.0260	0.5200	342.00	0.0200
7	0.18904	0.92927	0.20343	121.00	200.00	10.30	3.14	3375	28.80	0.0235	0.2150	309.00	0.0100
8	0.12596	0.60647	0.20768	254.00	350.00	12.50	4.32	6000	16.20	0.0340	0.4000	546.00	0.0200
9	0.16121	0.66724	0.24160	254.00	350.00	12.50	4.32	5500	15.70	0.0260	0.4100	566.00	0.0200
10	0.14256	0.75904	0.18781	121.00	232.50	9.70	4.42	3563	26.85	0.0290	0.5050	330.50	0.0150
11	0.14234	0.73331	0.19411	121.00	232.50	9.70	4.42	3688	26.70	0.0295	0.4300	331.00	0.0150
12	0.22893	0.74831	0.30593	121.00	200.00	9.60	4.00	3688	26.23	0.0135	0.5475	338.25	0.0150
13	0.13873	0.83151	0.16684	317.00	560.00	12.50	4.06	4000	14.50	0.0450	0.3300	618.50	0.0300
14	0.30841	0.84139	0.36654	195.00	265.00	12.50	3.81	4083	19.12	0.0235	0.1383	206.50	0.0100
15	0.10894	0.90266	0.12068	254.00	420.00	12.50	4.57	4000	15.35	0.0540	1.0500	577.50	0.0250
16	0.33618	0.78173	0.43004	121.00	200.00	9.60	3.14	3875	25.35	0.0125	0.1850	157.00	0.0100
17	0.14296	0.80642	0.17727	254.00	350.00	12.50	3.31	4500	20.10	0.0620	0.3600	443.00	0.0400
18	0.26674	0.85190	0.31311	254.00	354.00	12.50	3.70	4250	17.13	0.0173	0.4300	236.50	0.0100
19	0.14976	0.72198	0.20743	121.00	211.00	9.60	3.09	4250	24.80	0.0280	0.6800	357.00	0.0200
20	0.18805	0.71112	0.26444	121.00	200.00	10.50	3.54	4250	24.40	0.0180	0.4500	364.00	0.0100
21	0.13734	0.79375	0.17303	317.00	429.00	12.50	3.80	4750	17.20	0.0390	0.3300	516.00	0.0300
22	0.15905	0.72147	0.22045	254.00	344.00	11.00	3.66	4500	16.77	0.0240	0.4200	536.67	0.0267
23	0.15799	0.77302	0.20438	366.00	450.00	10.80	3.32	5000	15.90	0.0270	0.9500	555.00	0.0200

Notes: CID = cubic inch displacement, RHP = rated horsepower, cmp = compression ratio, axle = axle ratio, ETW = equivalent test weight (lb), MPG = fuel economy, HC = hydrocarbon emissions, CO = carbon monoxide emissions, CO_2 = carbon dioxide emissions, NOx = nitrogen oxide emissions

Table 10.7 Data and test results for selected carlines of European Company 2 (EC2)

Carline (DMU)	Overall Efficiency	Stage 1 Efficiency	Stage 2 Efficiency	CID	RHP	cmp	Axle	ETW	MPG	HC	CO	CO_2	NOx
1	0.39640	0.82057	0.48308	365.00	528.00	11.30	3.06	4750	14.60	0.0120	0.2500	610.00	0.0100
2	0.54335	0.89347	0.60814	213.00	268.00	10.70	2.82	3938	21.45	0.0075	0.1050	415.00	0.0100
3	0.14423	0.89662	0.16086	333.00	382.00	10.70	2.65	4250	18.00	0.0370	0.3200	496.00	0.0200
4	0.25589	0.81260	0.31490	213.00	268.00	10.50	3.07	4125	20.70	0.0150	0.2450	429.00	0.0100
5	0.18245	0.75571	0.24143	182.00	210.00	16.50	3.45	5500	22.50	0.0350	0.2500	452.00	0.0200
6	0.54032	0.86488	0.62473	379.00	507.00	11.30	2.82	4500	14.75	0.0090	0.1450	604.00	0.0100
7	0.18360	0.83993	0.21860	336.00	510.00	9.00	2.65	4750	14.10	0.0270	0.3700	629.00	0.0300
8	0.24238	0.85602	0.28315	365.00	603.00	9.00	2.65	4750	14.10	0.0220	0.6400	629.00	0.0100
9	0.14901	0.74522	0.19995	379.00	518.00	11.30	3.06	5000	13.80	0.0310	0.6100	643.00	0.0200
10	0.17705	0.79097	0.22383	336.00	510.00	9.00	2.65	5250	13.60	0.0280	0.4400	651.00	0.0400
11	0.21057	0.98959	0.21279	182.00	210.00	16.50	2.65	4250	29.10	0.0955	1.1250	348.00	0.0800
12	0.27448	0.87494	0.31372	183.00	228.00	11.30	3.07	3875	22.30	0.0140	0.2400	399.00	0.0200
13	0.60577	0.87150	0.69509	379.00	503.00	11.30	2.82	4750	15.50	0.0080	0.2000	576.00	0.0100
14	0.20322	0.87494	0.23227	183.00	228.00	11.30	3.07	3875	16.40	0.0230	0.1500	382.00	0.0200
15	0.25124	0.90030	0.27906	213.00	268.00	10.50	2.82	3875	21.30	0.0160	0.2500	417.00	0.0100
16	0.18004	0.77125	0.23345	333.00	382.00	10.70	2.65	5000	17.40	0.0260	0.2100	510.00	0.0200
17	0.48054	0.77398	0.62087	333.00	382.00	10.70	3.07	4500	16.50	0.0100	0.0500	535.00	0.0100
18	0.18584	0.58832	0.31589	213.00	268.00	10.50	3.90	5250	18.40	0.0190	0.2700	480.00	0.0200
19	0.77495	0.77495	1.00000	213.00	268.00	10.70	3.27	4250	20.30	0.0060	0.0200	435.00	0.0100
20	0.65856	0.85382	0.77131	379.00	450.00	11.30	2.82	4250	15.20	0.0070	0.2000	588.00	0.0100
21	0.16170	0.58670	0.27560	213.00	268.00	10.50	3.70	5500	18.20	0.0220	0.5000	485.00	0.0200
22	0.22994	0.95256	0.24139	333.00	382.00	10.70	2.47	4250	18.40	0.0210	0.2200	482.00	0.0200
23	0.45850	0.68762	0.66680	379.00	503.00	11.30	3.45	5500	12.80	0.0100	1.0300	696.00	0.0300

Notes: CID = cubic inch displacement, RHP = rated horsepower, cmp = compression ratio, axle = axle ratio, ETW = equivalent test weight (lb), MPG = fuel economy, HC = hydrocarbon emissions, CO = carbon monoxide emissions, CO_2 = carbon dioxide emissions, NOx = nitrogen oxide emissions

overall efficiency (0.86034) among 543 carlines in our test. Although this vehicle does not define the most efficient design in either Stage 1 or Stage 2, its overall efficiency is the highest as a result of the rather high design efficiencies in both the first and second stages (0.87879 and 0.97900). However, not all the hybrid vehicles perform well in our tests. For example, carline #11, a mid-size hybrid car produced by AC2 (Table 10.3), has a moderate first-stage efficiency (0.75497) but a low second-stage efficiency (0.25382), which leads to a relatively poor overall efficiency (0.19163).

For the two American manufacturers, carlines produced by AC1 perform generally well in the DEA test, as shown in Table 10.2. In particular, carline #4, a mid-size car powered by an internal combustion engine (ICE), has the highest overall efficiency (0.64428) among all the carlines produced by AC1 with relatively high efficiencies in both Stage 1 and Stage 2 (0.77061 and 0.83606). In addition, carline #8, a hybrid SUV produced by AC1, defines the efficient design of Stage 1 (first-stage efficiency = 1.0000) according to Table 10.2. In contrast, carlines produced by AC2 generally do not perform well in the DEA test, as shown in Table 10.3. Carline #3, a mid-size ICE car, and carline #13, a large-size ICE car, are two exceptions with relatively high overall efficiencies (0.61158 and 0.54485).

The two Japanese manufacturers perform moderately well in the DEA test. For JC1, in addition to the hybrid vehicle (carline #2) with the highest overall efficiency among all the carlines in our test, the carline with the second highly overall efficiency (0.55519) among all the vehicles produced by the company is carline #11, a hybrid mid-size car. Besides hybrid vehicles, carline #15, a compact ICE car, has the third highest overall efficiency (0.35105), as shown in Table 10.4. For JC2, carline #8 and #9, the sedan and coupe versions of a mid-size ICE car, has the highest and second highest overall efficiencies (0.51893 and 0.50382) among all the vehicles produced by the company, and carline #15, a compact ICE SUV, has the third highest overall efficiency (0.49248), as shown in Table 10.5.

For the two European manufacturers, carlines produced by EC1 perform generally poorly in the DEA test. Carline #4 and carline #5, the sedan and sport-wagon versions of a compact ICE car, have the first and second highest overall efficiencies (0.34554 and 0.32735) among vehicles produced by the company. Carline #14, another compact ICE car, has the third highest overall efficiency (0.30841) according to Table 10.6. In contrast, carlines produced by EC2 perform generally well in the test. In particular, carline #19, a large-size ICE car, not only defines the efficient design in Stage 2 (second-stage efficiency = 1.0000), but also has the highest overall efficiency (0.77495) among vehicles produced by the company. In addition, carline #20, a compact ICE car, has the second highest overall efficiency (0.65856) and relatively high efficiencies in both Stage 1 and Stage 2 (0.85382 and 0.77131), as shown in Table 10.7.

Based on the limited test results, we now present a number of interesting observations regarding sustainable product design. Technology innovation for expanding the efficient envelope/frontier, such as the development of hybrid technologies, is an important way for a firm to achieve high design efficiencies, as exemplified by the good industrial and bio design performances of some of the hybrid vehicles (e.g., carline #2 of JC1 and carline #8 of AC1) in our DEA test. However, high design efficiencies can also be achieved through innovative design choices to find the most efficient combination of product specifications and attributes that leads to high environmental performances even in the absence of advanced technologies, as exemplified by those ICE vehicles with high design efficiencies (e.g., carline #4 of AC1 and carline #19 of EC2). In many cases, finding the efficient combinations of product specifications/attributes may sometimes be an even more effective way to achieve higher design efficiencies than expanding the technology envelope/frontier through technology innovation. As shown in our test results, a large-size ICE car (carline #19 of EC2) may define the efficient design for the second stage, while a hybrid vehicle (carline #11 of AC2) may have poor design efficiencies. In fact, the test results suggest that, while the first-stage design efficiencies for industrial design of most carlines are reasonably high, there is still plenty of room for further improvement to enhance the second-stage efficiencies for bio design for most carlines and for most automobile manufacturers.

We now discuss how the proposed methodology can be used to improve decision-making in both the public and private sectors in the two applications presented previously. For Application #1, the comparative results given in Fig. 10.2 can be used by the environmental protection agency to understand the overall sustainable design efforts by different automakers in both the industrial design process and bio design processes. The performance comparison in Table 10.1 can also be used to evaluate the design efficiencies of different automakers, and adjust its regulatory approaches or policy instruments (e.g., taxes, subsidies, emissions standards, etc.) accordingly. For Application #2, the test results in Tables 10.1–10.7 can be used to benchmark the private company's design efficiencies, to conduct internal and external evaluations of different design options, as well as to identify the more eco-efficient ways to achieve environmental performances. Compared to the results in other DEA models for sustainable design (e.g., Linton 2002, Liu 2008, and Lin and Kremer 2010, which are all based on single-stage DEA), our test results provide a clearer picture of the individual efficiencies of both the industrial design process and sustainable design process as well as the overall sustainable design efficiency, which allows decision makers to better allocate their design efforts as well as to adjust the private strategies and public policies to induce more eco-efficient product designs.

It is noted that the DEA test can also be done by removing all the hybrid vehicles, but this is not likely to affect the test results because there exist ICE cars that define the efficient designs of Stage 1 and Stage 2, respectively, as the frontier units (e.g., carline #15 of AC2 and carline #19 of EC2). It should also be reiterated that our test results are limited by the assumption of using the centralized approach and by the incomplete product information provided in the database. With complete information of product specifications, attributes, and environmental performances as well as the exact design strategies (simultaneous, proactive, or reactive) adopted by firms, decision makers would be able to accurately assess the design performances through the network DEA model. It should also be noted that the dual model is not studied in the paper. While Kao and Hwang (2008) provide and discuss the dual model in the multiplier form, studying the dual model will not provide additional information related to the efficiency scores. If, however, assurance region (AR) type of information is available (Thompson et al. 1990), one would use the dual model to incorporate these AR constraints. The current paper does not have this type of information available. Thus, we leave this as a future topic for application.

10.6 Conclusion

In this paper, we propose a methodology with the use of two-stage network DEA for evaluating sustainable product design performances. We conceptualize design efficiency as a key measurement of design performance, and develop a network DEA model to link key engineering specifications, product attributes, and environmental performances in sustainable design. We also discuss how to use the centralized and decentralized models to analyze the simultaneous, proactive, and reactive approaches adopted by firms for sustainable design. In addition, we use data of engineering specifications, product attributes, and emissions performances in the vehicle emissions testing database to demonstrate the real-world applications of our DEA model for evaluating sustainable design performances in both the public and private sectors. The main message delivered here is that sustainable design does not need to mean compromise between traditional and environmental attributes. Through innovative design decisions for material selection, product reengineering, as well as expanding the technology envelope/frontier, a firm can find the most efficient way to combine product specifications and attributes which leads to better environmental performances. Our DEA-based methodology provides an innovative tool for decision makers to implement the win-win type of product design and innovation strategies for achieving the long-term sustainability of human society.

References

Atasu A, Souza GC (2010) How does product recovery affect quality choice? Working Paper. George Institute of Technology, Atlanta

Atasu A, Ozdemir O, Van Wassenhove LN (2008) Remanufacturing as a marketing strategy. Manag Sci 54(10):1731–1747

Calcott P, Walls M (2000) Can downstream waste disposal policies encourage upstream design for environment? Am Econ Rev 90(2):233–237

Cariaga I, El-Diraby T, Osman H (2007) Integrating value analysis and quality function deployment. J Constr Eng Manag 133(10):761–770

Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision making units. Eur J Oper Res 2:429–444

Chen C (2001) Design for the environment: a quality-based model for green product development. Manag Sci 47(2):250–263

Chen C, Zhang J (2009) The inconvenient truth about improving vehicle fuel efficiency: A multi-attributes analysis of the technology efficient frontier of the US automobile industry. Transport Res D 14(1):22–31

Chen Y, Cook WD, Li N, Zhu J (2009) Additive efficiency decomposition in two-stage DEA. Eur J Oper Res 196:1170–1176

Chiou YC, Lan LW, Yen TH (2010) A joint measurement of efficiency and effectiveness for non-storable commodities: integrated data envelopment analysis approaches. Eur J Oper Res 201(2):477–489

CNTV (2010) New car emission standard imposed. CNTV (June 12)

Conway-Schempf N, Lave LB (1999) Green design tools for environmental management. Environ Qual Manag 8(4):35–46

Cook WD, Liang L, Zhu J (2010) Measuring performance of two-stage network structures by DEA: a review and future perspective. OMEGA 38:423–430

Cooper WW, Seiford LM, Tone K (2007) Data envelopment analysis: a comprehensive text with models, applications, references and DEA-solver software, 2nd edn. Springer, New York

Crandall RW, Graham JD (1989) The effect of fuel economy standards on automobile safety. J Law Econ 32(1):97–118

Dekker R, Fleischmann M, Inderfurth K, Van Wassenhove LN (2010) Reverse logistics: quantitative models for closed-loop supply chains. Springer, New York

Design News (2008) The fuel-efficient automotive transmissions of tomorrow. Design News 63(9) (Jun 23)

Dyson RG, Allen R, Camanho AS, Podinovski VV, Sarrico CS (2001) Pitfalls and protocols in DEA. Eur J Oper Res 132:245–259

Emrouznejad A, Parker BR, Tavares G (2010) Evaluation of research in efficiency and productivity: a survey and analysis of the first 30 years of scholarly literature in DEA. Socioecon Plann Sci 42(3):151–157

U.S. EPA (2009) Certification emission test data. Available at www.epa.gov

U.S. EPA (2010) EPA and NHTSA finalize historic national program to reduce greenhouse gases and improve fuel economy for cars and trucks. EPA-420-F-10-014 (April 2010)

European Federation for Transport and Environment (2008) Weight-based standards make CO_2 targets harder to reach. Background Briefing (April 2008)

Färe R, Grosskopf S (1996) Productivity and intermediate products: A frontier approach. Econ Lett 50:65–70

Färe R, Grosskopf S (2000) Network DEA. Socioecon Plann Sci 34:35–49

Farrell MJ (1957) The measurement of production efficiency. J Roy Stat Soc Ser A 120(3):253–81

Farris JA, Groesbeck RL, Van Aken EM, Letens G (2006) Evaluating the relative performance of engineering design projects: a case study using data envelopment analysis. IEEE Trans Eng Manag 53(3):472–482

Fiksel J (2009) Design for environment: A guide to sustainable product development. McGraw-Hill

Fleischmann M, Bloemhof-Ruwaard JM, Dekker R, van der Laan E, van Nunen JAEE, Van Wassenhove LN (1997) Quantitative models for reverse logistics: a review. Eur J Oper Res 103:1–17

Fullerton D, Wu W (1998) Policies for green design. J Environ Econ Manag 36:131–148

Graedel TEH, Allenby BR (2009) Industrial ecology and sustainable engineering. Prentice Hall

Green Brand Survey, 2012. WPP Companies Cohn & Wolfe, Landor Associates, and Penn Schoen Berland

Handfield RB, Melnyk SA, Calantone RJ, Curkovic S (2001) Integrating environmental concerns into the design process: the gap between theory and practice. IEEE Trans Eng Manag 48(2):189–208

Hauser JT, Clausing D (1988) The house of quality. Harv Bus Rev 66(3):63–73

Hendrickson CT, Lave LB, Matthews HS (2006) environmental life cycle assessment of goods and services: an input-output approach. RFF, Washington, DC

Hopkins MS (2010) How sustainability fuels design innovation. Sloan Manage Rev 52(1):75–81
Kao C, Hwang SN (2008) Efficiency decomposition in two-stage data envelopment analysis: an application to non-life insurance companies in Taiwan. Eur J Oper Res 185(1):418–429
Lauridsen EH, Jørgensen U (2010) Sustainable transition of electronic products through waste policy. Res Policy 39(4):486–494
Liang L, Cook WD, Zhu J (2008a) DEA models for two-stage processes: game approach and efficiency decomposition. Nav Res Logist 55:643–653
Liang L, Wu J, Cook WD, Zhu J (2008b) The DEA game cross efficiency model and its Nash equilibrium. Oper Res 56:1278–1288
Lin C-Y, Kremer GEO (2010) DEA applications in the product design domain. VDM, Saarbrücken
Linton JD (2002) DEA: a method for ranking the greenness of design decisions. J Mech Des 124:145–150
Malloy RA (1996) Designing with recyclate. Plastics World 54:27–28
Miyashita T (2000) A study on the collaborative design using supervisor system. Trans Jpn Soci Mech Eng 66(643):921–928
Noori H (1990) Managing the dynamics of new technology. Prentice Hall, Englewood Cliffs
Noori H, Chen C (2003) Applying scenario-driven strategy to integrate environmental management and product design. Prod Oper Manag 12(3):353–368
Shephard RW, Färe R (1979) Dynamic theory of production correspondences. Anton Hain, Meisenheim
Thompson RG, Langemeier LN, Lee CT, Lee E, Thrall RM (1990) The role of multiplier bounds in efficiency analysis with application to Kansas farming. J Econ 46:93–108
Ulrich KT, Eppinger SD (2012) Product design and development. McGraw-Hill, New York
Urban GL, Hauser JR (1993) Design and marketing of new products. Prentice Hall, Upper Saddle River
Verhoef EV, Dijkema GPJ, Reuter MA (2004) Process knowledge, system dynamics, and metal ecology. J Ind Ecol 8(1-2):23–43

Chapter 11
Measuring Environmental Efficiency: An Application to U.S. Electric Utilities

Chien-Ming Chen and Sheng Ang

Abstract This chapter highlights limitations of some DEA (data envelopment analysis) environmental efficiency models, including directional distance function and radial efficiency models, under weak disposability assumption and various return-to-scale technology. It is found that (1) these models are not monotonic in undesirable outputs (i.e., a firm's efficiency score may increase when polluting more, and vice versa), (2) strongly dominated firms may appear efficient, and (3) some firms' projection points derived from the optimal environmental efficiency scores are strongly dominated, thus they cannot be the right direction for the improvement. To address these problems, we propose a weighted additive model, i.e., the Median Adjusted Measure (MAM) model. An application to measuring the environmental efficiency of 94 U.S. electric utilities is presented to illustrate the problems and to compare the existing models with our MAM model. The empirical results show that the directional distance function and radial efficiency models may generate spurious efficiency estimates, and thus it must be with caution.

Keywords Data envelopment analysis • Environmental efficiency • Undesirable outputs • Various return-to-scale • Electric utilities

This chapter is prepared based on Chen (2013, 2014). The authors thank Springer and Elsevier for granting the rights for reusing the contents of these two papers. The authors also thank Maria Montes-Sancho for preparing the data used in this article.

C.-M. Chen (✉)
Nanyang Business School, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798, Singapore
e-mail: cmchen@ntu.edu.sg

S. Ang
School of Management, University of Science and Technology of China, 96 Jinzhai Road, Hefei, Anhui 230026, PR China
e-mail: shengang@ustc.edu.cn

J. Zhu (ed.), *Data Envelopment Analysis*, International Series in Operations Research & Management Science 238, DOI 10.1007/978-1-4899-7684-0_11

11.1 Introduction

When measuring environmental efficiency, we seek to answer the following question: Can a firm produce more desirable outputs while generating lower quantities of undesirable outputs than its competitors? The answer to this question can help managers and policymakers act pro-actively in strategy-making and resource allocation to ensure both corporate and environmental sustainability. However, measuring environmental efficiency can be challenging for several reasons. First, calculating environmental efficiency scores requires an articulation of weights or preferences for productive inputs and outputs, but both eliciting and combining preferences are difficult in a multi-stakeholder environment (Baucells and Sarin 2003). Second, most undesirable outputs, such as greenhouse gas emissions and toxic releases, do not have a well-established market from which we can obtain reliable price signals. This makes prioritizing different environmental factors difficult. For example, it can be difficult to assign specific weights to different dimensions of corporate social performance, such as environmental consciousness and community relationship (Chen and Delmas 2011).

The absence of reliable price information for environmental impacts makes data envelopment analysis (DEA) a useful tool for assessing environmental efficiency. DEA does not require explicit assumptions about weights, production functions, and probability distributions for environmental inefficiency. Weights are optimized based on which input(s) a specific firm excels at utilizing, or which output(s) a firm excels at generating in comparison to the other firms in the sample. In this way, each firm can endogenously determine the weights used to evaluate its eco-efficiency. Applications of DEA to environmental efficiency have also been in a variety of problem contexts where undesirable outputs are consequential, including banking and finance, electricity generation, manufacturing, and transportation. The goal of this chapter is to review the commonly used DEA models for measuring environmental efficiency and talk about their potential limitations.

In the DEA literature, the directional distance function (DDF) (Chung et al. 1997) and radial efficiency models (e.g., Zhou et al. 2007; Färe et al. 1989) are among the two most widely used. Compared with other DEA models (e.g., Seiford and Zhu 2002), the DDF and radial efficiency models usually adopt an additional assumption on undesirable outputs, i.e., weak disposability assumption (WDA) on undesirable outputs (Shephard 1970), which specifies the trade-off (and boundary) relationship between a firm's capability to produce good and bad outputs in the production possibility set. This chapter reveals three problems associated with these two models under the weak disposability assumption: (1) - Non-monotonicity in undesirable outputs: a firm's efficiency obtained from the two models may increase when polluting more, and vice versa, (2) misclassification of efficiency status: strongly dominated firms may be identified efficient, and (3) strongly dominated projection targets: environmental efficiency scores may be computed against strongly dominated points. Our findings suggest that the DDF and radial efficiency models should be used with caution. We also examine modelling issues under variable returns-to-scale (VRS) production technology. As a solution, we propose an alternative model based on the weighted additive model (Cooper et al. 1999), and compare our model with the existing models by an illustrative application of evaluating the environmental efficiency of 94 U.S. electric utilities in year 2007.

In the next section, we introduce the production technology assumptions, the DDF and radial efficiency models for environmental efficiency evaluation, and identified issues and problems. In Sect. 11.3, we develop a model to avoid the problems of the existing models. In Sect. 11.4 we include a case study for measuring the environmental efficiency of 94 U.S. electric utilities. Section 11.5 gives conclusions.

11.2 Production Models with Undesirable Outputs for Environmental Efficiency

11.2.1 Production Technology Assumptions

We consider n decision-making units (DMU). Each DMU uses m inputs to produce s desirable outputs and p undesirable outputs. The input vector of DMU q is denoted by $X_q = \left(x_{q1}, \ldots, x_{qm}\right)$, desirable output vector by $Y_q = \left(y_{q1}, \ldots, y_{qs}\right)$, and undesirable output vector by $B_q = \left(b_{q1}, \ldots, b_{qp}\right)$. The correspondence between the three vectors can be described as:

$$f\left(X_q\right) \triangleq \left\{\left(Y_q, B_q\right) : \left(Y_q, B_q\right) \text{ can be produced by using } X_q\right\}. \tag{11.1}$$

The function f captures the relationship between inputs and outputs and hence represents the production technology. A common behavioural assumption is that producer q should maximize Y_q and minimize B_q for a given X_q. We define output efficiency as:

Definition 1 (output efficiency) *DMU q is output efficient if there does not exist a non-zero vector $\left(S^Y, S^B\right) \in \mathfrak{R}_+^s \times \mathfrak{R}_+^p$, such that $\left(Y_q + S^Y, B_q - S^B\right) \in f\left(X_q\right)$.*

Definition 1 means that a DMU is output efficient if it is impossible to improve any of its outputs given the current input level. Note that output efficiency defined here is similar to but different from the Pareto-Koopmans efficiency (Cooper et al. 2007, pp. 45–46), in that output efficiency does not consider input-side inefficiency and slacks (i.e., reductions in some of the inputs).

The definition of output efficiency implies that firms can improve output efficiency by either increasing Y_q, decreasing B_q, or both. This entails the question of how to model the trade-off relationship between the desirable and undesirable outputs. One possibility is to assume there is no such trade-off, *ceteris paribus*. In this situation (i.e., *free disposability*), the technology set (X,$f(X)$) allows lowering undesirable outputs without losing desirable outputs; i.e., $\left(Y_q, B_q\right) \in f\left(X_q\right) \Rightarrow \left(Y_q, B_q^*\right) \in f\left(X_q\right)$ for $B_q^* \geqq B_q$, and $\left(Y_q, B_q\right) \in f\left(X_q\right) \Rightarrow \left(Y_q^*, B_q\right) \in f\left(X_q\right)$, for all $Y_q^* \leqq Y_q$ and $Y_q^* \in \mathfrak{R}_+^s$, "$\leqq$" being the component-wise inequality. Alternatively, one may assume reducing undesirable outputs should not be "free" and impose a weak disposability assumption on undesirable outputs. Denoting the technology set under the weak disposability assumption as $f_w(X_q)$, the weak disposability assumption satisfies the following three conditions (Shephard 1970): (i) $\left(Y_q, B_q\right) \in f_w\left(X_q\right)$ implies that $\left(Y_q^*, B_q\right) \in f\left(X_q\right)$ for all $Y_q^* \leqq Y_q$, (ii) $\left(Y_q, B_q\right) \in f_w\left(X_q\right)$ and $0 \leq \theta \leq 1$ implies that $\left(\theta Y_q, \theta B_q\right) \in f_w\left(X_q\right)$, and (iii) $\left(Y_q, B_q\right) \in f_w\left(X_q\right)$ implies that $\left(Y_q, B_q\right) \in f\left(X_q^*\right)$ for all $X_q^* \geqq X_q$.

The first condition means that if (X_q, Y_q, B_q) is observed, the existence of this observation implies that it is feasible to produce a lower amount of desirable outputs with given X_q and B_q. The second condition stipulates that proportional reduction of the joint output vector (Y_q, B_q) is feasible. The first two conditions imply that a reduction in B_q must be accompanied by a reduction in desirable outputs Y_q, while the converse is not true. The weak disposability assumption condition is meant to reflect that generation and disposal of undesirable outputs should not be free, in a sense that reducing undesirable outputs will come at the expense of lowering desirable outputs. Clearly, the technology set $f_w(X_q)$ is a subset of $f(X_q)$, because of these additional constraints associated with the weak disposability assumption.

The technology f_w can be formulated as a linear system under the following axioms: $f_w(X_q)$ is convex, and $f_w(X_q)$ is the intersection of all sets satisfying the convexity axiom and disposability assumptions; i.e., the production set $f_w = \bigcap_{j=1}^{n} f'_w\left(X_q\right)$, where $f'_w(X_q)$ is any convex set satisfying the disposability assumption for DMU j (Banker et al. 1984). The model can be expressed as:

$$f_w\left(X_q\right) = \left\{ (Y, B) : \sum_{j=1}^{n} \lambda_j x_{ji} \leq x_{qi}, i = 1, \ldots, m, \sum_{j=1}^{n} \lambda_j y_{jr} \geq y_{qr}, r = 1, \ldots, s, \right.$$
$$\left. \sum_{j=1}^{n} \lambda_j b_{jk} = b_{qk}, k = 1, \ldots, p, \lambda_j \geq 0, j = 1, \ldots, n \right\} \tag{11.2}$$

The boundary of (11.2) consists of non-negative linear combinations of all DMUs' input and output vectors. The λ_j represents the production intensity of the jth DMU, which can take different values to populate different areas of $(X_q, f_w(X_q))$. The weak disposability assumption is enforced by the equality constraints associated with undesirable outputs. See p. 50 in Färe and Grosskopf (2006) for the proof that shows (11.2) satisfies the weak disposability assumption. If on contrary we assume that undesirable outputs are freely disposable, the new technology set $f_f(X_q)$ can be recast by replacing the equality constraints with "$\leq$" inequality constraints, meaning that the efficient level of undesirable outputs are bounded below by the left-hand-side value and undesirable outputs can be improved independently from desirable outputs.

Note that the convex set $(X_q, f_w(X_q))$ satisfies the constant returns-to-scale (CRS) assumption; i.e., $(Y, B) \in f_w(X)$ implies that $(\delta Y, \delta B) \in f_w(\delta X), \delta \geq 0$. A number of studies on environmental efficiencies assume a VRS technology (Chen 2013). These studies follow Banker et al. (1984) and add a convexity constraint on the intensity variables to represent the VRS assumption imposed (e.g., Mandal and Madheswaran 2010; Oggioni et al. 2011; Riccardi et al. 2012). However, it is a general misconception that simply adding a convexity constraint to the CRS model with weak disposability means that the new model is one with a VRS technology with weak disposability, as shown in Färe and Grosskopf (2003). As such, many studies used an incorrect VRS formulation in the literature (Chen 2013).

The correct VRS formulation with weak disposability assumption first appeared in Shephard (1970). However, the Shephard's VRS formulation with weak disposability is highly nonlinear and thus the model has difficulties in computation. Also the production set under the Shephard's VRS formulation is not convex, which means that some of the feasible points in the production set under the convexity axiom in nonparametric production models (see, e.g., Banker et al. 1984) may be deemed infeasible in Shephard's formulation. Kuosmanen (2005) and Kuosmanen and Podinovski (2009) extend Shephard's VRS formulation by developing a convex and fully linearizable model (i.e., linearizable for all common types of efficiency indexes):

$$f_{VRS}(X_q) = \left\{ (Y, B) : \sum_{j=1}^{n} (\lambda_j + \mu_j) x_{ji} \leq x_{qi}, i = 1, \ldots, m; \sum_{j=1}^{n} \lambda_j y_{jr} \geq y_{qr}, r = 1, \ldots, s; \right.$$
$$\left. \sum_{j=1}^{n} \lambda_j b_{jk} = b_{qk}, k = 1, \ldots, p; \sum_{j=1}^{n} (\lambda_j + \mu_j) = 1; \lambda_j, \mu_j \geq 0, j = 1, \ldots, n \right\} \tag{11.3}$$

It is shown that the Shephard's VRS formulation is a special case of the Kuomanen's VRS formulation (Kuosmanen 2005). More importantly, the efficiency models constructed based on (11.3) become linear programming problems and can be solved easily. However, to date few papers in the literature have employed this general and correct VRS formulation in environmental efficiency analysis (Chen 2013). Next, we introduce the DDF and radial efficiency models based on Kuosmanen's formulation.

11.2.2 Directional Distance Function

The formulation of the directional distance function (DDF) is shown in (11.4). Specifically, the DDF model calculates the environmental efficiency score of a firm according to the maximum improvement in outputs that this firm can make in the direction (g^Y, g^B), such that the firm remains in $f_{VRS}(X_q)$ after this improvement. Therefore environmentally efficient firms in the DDF model are those obtaining a zero optimal value (i.e., $\theta^* = 0$), in a sense that these firms cannot improve their outputs following the pre-determined direction.

$$\begin{aligned}
&\text{Max}\theta\\
&\text{s.t.}\sum_{j=1}^{n}(\lambda_j+\mu_j)x_{ji}\le x_{qi}, i=1,\ldots,m\\
&\sum_{j=1}^{n}\lambda_j y_{jr}\ge y_{qr}+\theta g_r^Y, r=1,\ldots,s\\
&\sum_{j=1}^{n}\lambda_j b_{jk}= b_{qk}-\theta g_k^B, k=1,\ldots,p\\
&\sum_{j=1}^{n}(\lambda_j+\mu_j)=1\\
&\lambda_j,\mu_j\ge 0, j=1,\ldots,n
\end{aligned}\tag{11.4}$$

We can calculate the projection point for each DMU according to the efficiency score obtained from (11.4). For example, $\left(X_q, Y_q+\theta^* g^Y, B_q-\theta^* g^B\right)$ is the projection point of DMU q under DDF, where θ^* is the optimal solutions to the corresponding efficiency model (11.4). Clearly, the projection point is at the boundary of the production set. As noted, the projection point is the linear combination of different observed DMUs. We define *the reference set* for *an evaluated DMU* as the collection of DMUs that forms the projection point. The λ's associated with these active DMUs are positive in the optimal solution (Cooper et al. 2007). Thus this also means that an efficient DMU is its own reference set and projection point.

11.2.3 Radial Efficiency Models

Studies with a radial efficiency index (Charnes et al. 1978; Farrell 1957) under the weak disposability assumption are found in the literature. The number of papers using radial efficiency models increases rapidly over past years (Chen 2013). These models with a radial efficiency index can be classified into the follow three types[1]: the index associated with desirable outputs and undesirable outputs, desirable outputs only, and undesirable outputs only, which can be modelled by (11.5).

$$\begin{aligned}
&\text{Max}\,\theta \quad \left(or\ \text{Min}\,\delta^b\right) \\
&\text{s.t.}\sum_{j=1}^{n}\left(\lambda_j+\mu_j\right)x_{ji} \leq x_{qi}, i=1,\ldots,m \\
&\sum_{j=1}^{n}\lambda_j y_{jr} \geq \theta y_{qr}, r=1,\ldots,s \\
&\sum_{j=1}^{n}\lambda_j b_{jk} = \delta^b b_{qk}, k=1,\ldots,p \\
&\sum_{j=1}^{n}\left(\lambda_j+\mu_j\right)=1 \\
&\lambda_j,\mu_j \geq 0, j=1,\ldots,n
\end{aligned} \tag{11.5}$$

Before illustrate the limitations associated with the DDF and radial efficiency models in the next section, we list the four types of environmental efficiency models introduced thus for (M1 to M4 in Table 11.1). It should be noted that DMUs' efficiency scores obtained from models M1 to M4 have different ranges and different values for efficient observations. To be specific, a DMU having lower score in M1, M2 and M4 is considered more efficient, but M1 is equal or greater than zero while M2 and M4 have a lower bound of one. DMUs obtaining higher scores in M3 are considered more efficient and the range of M3 is from zero to one.

Table 11.1 Efficiency models classification for measuring environmental efficiency (Chen 2013)

	Models, or efficiency indexes associated with	Objective function in (11.5)	Modification in (11.5)	Range of efficiency score	Score of efficient observations
M1	Model (11.4); directional distance function	–	–	$[0,\infty]$	0
M2	Desirable outputs only	Maxθ	$\delta^b=1$	$[1\ \ \infty)$	1
M3	Undesirable outputs only	Minδ^b	$\theta=1$	(0, 1]	1
M4	Desirable outputs and undesirable outputs	Maxθ	$\delta^b=1/\theta$	$[1\ \ \infty)$	1

[1] There exists the fourth type: radial efficiency index attached with inputs only. But it is not presented here as we focus on output-oriented models in this chapter.

The projection point for DMU q according to the efficiency score and optimal solutions obtained from radial efficiency models is $(X_q, \theta^*Y_q, \delta^{b*}B_q)$, which is at the boundary of the production set.

11.2.4 Problems Illustration by a Numerical Example

We present a simple numerical example to show problems of the DDF and radial efficiency models with the WDA and Kuosmanen's VRS assumptions.

In this numerical sample, there are four observed DMUs (DMU A to D) with one input, one desirable, and one undesirable output, shown in Table 11.2. For the ease graphical presentation, all four DMUs are assumed to consume the same amount of inputs. The output set $f_{VRS}(X)$ for this sample based on the production technology model (11.3) is represented by the region '0ABCE0' in Fig. 11.1. When WDA is not imposed, the output set expands and becomes the area under the line segment '0A'

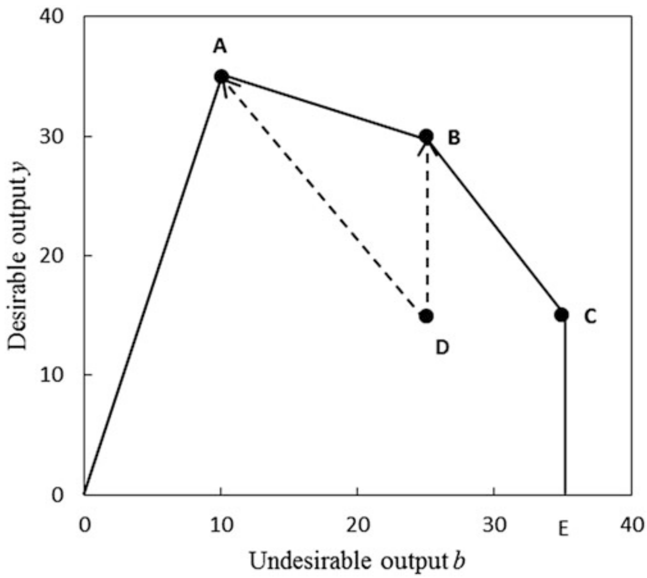


Fig. 11.1 Output set f_{VRS} under the WDA and Kuosmanen's VRS technology

Table 11.2 A numerical example for problems illustration

DMU	Input x	Undesirable output b	Desirable output y
A	10	10	35
B	10	25	30
C	10	35	15
D	10	25	15

and the horizontal line extended from A to its right. More specifically, for the desirable output y, which is freely disposable, the area below the line segments '0A' are considered feasible (c.f. the inequality constraint for y in (11.3)). Observe that the frontier under the WDA (i.e., the boundary of $f_{VRS}(X)$) may include points dominated in both y and b, which correspond to the problem of *misclassification of efficiency status*. For example, DMUs B and C produce a lower amount of y but more b than DMU A. However, DMUs B and C are in the boundary set of $f_{VRS}(X)$. DMU D may be projected to the dominated portion of the boundary set (i.e., the line segment between A and B) with certain choices of directional vectors. The same thing may occur in the radial efficiency models (e.g., M2 projects D to B, and M4 may project D to the line segment between A and B by a hyperbolical locus). This potential problem for DMU D is called the problem of *strongly dominated projection targets*. If we increase the undesirable output of D from 25 to 35, the inefficient DMU D would become efficient DMU C. That is to say, an increase in a DMU's undesirable outputs may improve the DMU's efficiency score, which correspond to the problem of *non-monotonicity in undesirable outputs*.

Chen (2014) proves the above three problems associated with the DDF and HEM (M4) models under CRS technology. In the next section, we introduce a weighted additive model for environmental efficiency evaluation as a solution.

11.3 A Median Adjusted Measure (MAM) Model for Environmental Efficiency

As noted earlier, weighted additive models have been shown to be able to project all DMUs onto the efficient facet (Charnes et al. 1985), which resolves the dilemma of choosing between free and weak disposability for undesirable outputs. Weighted additive models are a general class of models that include many variants (Charnes et al. 1985; Seiford and Zhu 2005; Färe and Grosskopf 2010). One important issue for implementing the weighted additive model is that we must specify weights. This is particular a problem as DEA models are known as a weight-free approach and do not require subjective weight assignments. Chen and Delmas (2012) use the DMU's own outputs to normalize the output improvements and then calculate environmental efficiency as the average normalized score. This approach has a potential limitation in that different DMUs would be based its own production but miss information about distributions of different outputs across the entire sample, which may carry significant practical implications. Some studies assign weights based on the sample statistics, such as the range adjusted measure (RAM) model proposed by Cooper et al. (1999):

$$
\begin{aligned}
&\text{Max}\Gamma = \frac{1}{s+p}\left(\sum_{r=1}^{s}\frac{s_r^+}{R_r^+} + \sum_{k=1}^{p}\frac{s_k^-}{R_k^-}\right)\\
&\text{s.t.}\sum_{j=1}^{n}(\lambda_j+\mu_j)x_{ji} \le x_{qi}, i = 1, \ldots, m\\
&\sum_{j=1}^{n}\lambda_j y_{jr} = y_{qr} + s_r^+, r = 1, \ldots, s\\
&\sum_{j=1}^{n}\lambda_j b_{jk} = b_{qk} - s_k^-, k = 1, \ldots, p\\
&\sum_{j=1}^{n}(\lambda_j+\mu_j) = 1; \lambda_j, \mu_j \ge 0, j = 1, \ldots, n\\
&\lambda_j, \mu_j \ge 0\ , j = 1, \ldots, n\\
&s_r^+, s_k^- \ge 0, r = 1, \ldots, s; k = 1, \ldots, p
\end{aligned}
\tag{11.6}
$$

where R_r^+ is the range of the rth desirable output and R_k^- is the range of the pth undesirable output. Note that the RAM model can also incorporate slacks variables for inputs. When the inputs slacks are taken into consideration, we need to replace the objective function in (11.6) by $\text{Max}\frac{1}{m+s+p}\left(\sum_{i=1}^{m}\frac{s_i^-}{R_i^-} + \sum_{r=1}^{s}\frac{s_r^+}{R_r^+} + \sum_{k=1}^{p}\frac{s_k^-}{R_k^-}\right)$ where R_i^- is the range of the ith input, and change the input inequality constraints in (11.6) to $\sum_{j=1}^{n}(\lambda_j+\mu_j)\, x_{ji} = x_{qi} - s_i^-, i = 1, \ldots, m$. For the purpose of the current paper, we focus on the output-oriented RAM model. For the economic intuition behind the RAM model, see Cooper et al. (1999) for an excellent exposition of the rationale behind the additive efficiency model and its use to measure allocative, technical, and overall inefficiencies.

We propose a model based on the concept from the RAM model, as Cooper et al. (1999) point out that the RAM-type of efficiency models come with a number of desirable properties, including (i) the efficiency score is bounded in [0,1], (ii) the model is unit invariant, (iii) the model is strongly monotonic in slacks, and (iv) the model is translation invariant under the variable returns-to-scale technological assumption (Banker et al. 1984). However, we find using ranges as the normalizing factors problematic, and choose to use other normalizing variables instead of ranges in the original model. For example, it is stated in Cooper et al. (1999) that $0 \le \Gamma \le 1$, where a zero value indicates efficiency and a value of one indicates full efficiency. As the slacks are usually much lower in magnitude than their corresponding ranges, the efficiency scores obtained from the original RAM model tends to be low in both magnitude and variation (Cooper et al. 1999; Steinmann and Zweifel 2001). Therefore the RAM scores cannot effectively differentiate the performance of different DMUs. Furthermore, if we observe extremely inefficient firms that makes certain R_r^+ and/or R_k^- larger. These extremely inefficient firms may be those that produce lower than minimal observed desirable outputs but higher than maximum observed undesirable outputs at a fixed input level. The efficiency scores of all the other firms may decrease markedly, and most firms would appear more efficient although the efficient frontier remains unaltered. As it is not uncommon to observe "heavy polluters" in applications, using ranges or other dispersion measures of outputs do not seem appropriate. Also note that if a weighted additive model is used, the disposability assumption on undesirable outputs will not have any impact on the resultant efficiency scores.

Another problem of using ranges is that ranges cannot reveal the relative magnitude of the output. For example, suppose we obtain for a particular DMU that its slack for an output is 5 and the corresponding range for that output is 50. The managerial implication of this output slack for this DMU may be quite different if the maximum and minimum of the output are respectively 10 and 60 rather than 500 and 550, for example. As the main purpose of the normalizing factors are to obtain unit invariance, we opt for using the median of outputs to replace the range used in the objective function of model (11.6), which is more robust than ranges or averages as the basic statistical properties of these measures. We call our efficiency measure based on median the "Median Adjusted Measure" (MAM). The MAM score then has an intuitive interpretation as the average of slacks compared to the sample median of the corresponding output variables. Note that one may designate the normalizing parameters in the original range adjusted model in other ways; see, e.g., Cooper et al. (2011) for a comprehensive discussion.

11.4 An Application to Measuring Environmental Efficiency of U.S. Electric Utilities

The electricity sector has been under stringent scrutiny for its environmental performance (Majumdar and Marcus 2001; Fabrizio et al. 2007; Delmas et al. 2007). Following previous studies (e.g., Majumdar and Marcus 2001; Delmas et al. 2007), we consider plant value, total operation & maintenance expenditure, labor cost, and electricity purchased from other firms as four input variables. The desirable output considered is total sales in MWH, and three undesirable outputs are sulfur dioxide (SO_2), nitrogen oxide (NO_x), and carbon dioxide (CO_2), of which

SO_2 and NO_x are regulated by the U.S. Environmental Protection Agency (EPA) under the Acid Rain Program.

The data are collected from the U.S. Federal Energy Regulatory Commission (FERC) Form Number 1 (U.S. DOE, FERC Form 1), from the U.S. Energy Information Administration (Forms EIA-860, EIA-861, and EIA-906), and from the U.S. Environmental Protection Agency Clean Air Market Program's website. Our sample consists of 94 major investor-owned electric utilities in 2007. Table 11.3 reports the statistics summary of the electric utilities' input, desirable and undesirable outputs, which show that the 94 utilities vary significantly in their production scales, thus a VRS technology assumption is employed to reflect the industry production technology. In the application of the 94 U.S. electric utilities, we apply the DDF and radial efficiency models to show the limitations under the WDA and VRS technologies, and also apply our proposed median adjusted measures model as an illustration.

We applied the models M1–M4. For DDF (M1), an all-one vector is employed as the directional vector which is fixed. Another commonly used directional vectors includes $g_q = \left(Y_q, B_q\right)$, $(0, B_q)$, or $(Y_q, 0)$, or sample average values of outputs. Although not shown here, the three problems mentioned in the previous sections will still persist under these alternative directional vectors.

Table 11.4 shows the environmental efficiency results and optimal slacks values from the MAM models. There are 17 firms identified as strongly efficient by the MAM model, because they have zero optimal slacks in both desirable and undesirable outputs and are efficient across all of M1 to M4 models at the same time. However, some of the firms appear efficient in models M1 to M4 are strongly dominated in their outputs such as firms #2, #5, and #17. The rate of misclassification is rather high for the DDF and radial efficiency models (average higher than 30 %).

We have obtained the optimal efficiency scores of all the firms by models M1 to M4 and the efficiency classification. These optimal efficiency scores can also be used to compute the projection points for those inefficient firms. To examine the problem of strongly dominated projection targets, we add the obtained projection points into the original data set, and use the MAM model to evaluate the efficiency of the projection points. Table 11.5 shows the efficiency results of those firms' projection points under models M1 to M4. Besides those output efficient firms under MAM, the efficiency scores of the other firms' projection targets under M1, M2, M3 and M4 are larger than zero (except DMU #57 under M1, DMU #27 and #87 under M3, and DMU #14 under M3). Thus, those projection targets are not strongly efficient and some are not even efficient in a weak sense.

Table 11.3 Variables and data set descriptive statistics (n=94)

Variable	Median	Max	Min	Std. Dev.
Inputs (in dollars)				
Plant value	3.8E + 09	4.21E + 10	1.65E + 08	7.01E + 09
Labor cost	70048960	9.41E + 08	1450388	1.9E + 08
Total operation and maintenance expenditure	9.24E + 08	8.23E + 09	55344397	1.47E + 09
Electricity purchased (MWH)	6602986	46016389	86797	8304988
Undesirable outputs (in tons)				
SO_2	39570.45	682271	7.108	112083.7
NO_X	20312	139549.9	47.10725	29908.47
CO_2	12945367	85832369	51480.4	20678124
Desirable output				
Total sales (MWH)	3591905	32168051	626	5969309

Table 11.4 Environmental efficiency scores and optimal slack values

DMU #	MAM	M1 (DDF)	M2	M3	M4 (HEM)	Optimal slacks of the additive model (MAM)			
						SO_2	NO_X	CO_2	Total sales
1	3.51	4983.03	1.07	0.36	1.08	298994.4	61235.6	45023463.3	0
2	1.07	0	1	1	1	38797.4	34159.3	21137806.7	0
3	0.74	0	1	0.20	2.01	1085.4	6593.4	3305457.4	8412368.9
4	0.59	3016.06	1.60	0.35	1.44	22927.6	5747.4	4477973.7	4065339.0
5	0.09	0	1	1	1	1332.0	313.4	1328602.7	766215.7
6	0	0	1	1	1	0	0	0	0
7	0.66	0	1	0.00*	192.83	9379.4	0	1901884.5	8118703.7
8	0.83	3841.68	10.18	0.02	5.08	11986.3	3201.1	1529910.1	9849195.5
9	0	0	1	1	1	0	0	0	0
10	0	0	1	1	1	0	0	0	0
11	2.31	15441.27	1.43	0.27	1.41	178519.1	50465.4	28941495.4	0
12	2.12	596.94	2.54	0.01	3.66	114655.7	22760.8	17959793.3	11095584.0
13	1.56	4826.41	1.39	0.50	1.28	130551.7	25185.8	12192186.3	2754306.0
14	0.63	938.37	10.70	0.06	3.72	0	0	418381.6	8959201.8
15	2.98	12936.85	10.57	0.00	6.45	149871.7	40557.2	35669394.3	12075297.0
16	2.05	20595.79	1.95	0.20	1.57	157706.0	46169.9	25085423.2	0
17	0.43	0	1	1	1	9287.1	0	92155.4	5255681.1
18	5.74	0	1	0.03	2.84	406933.9	109055.0	69836313.4	6849289.2
19	2.46	0	1	1	1	225305.4	24355.2	38287896.6	0
20	2.28	0	1	0.00*	50.96	167061.8	25536.7	12916281.1	9559643.5
21	0.21	0	1	1	1	0	3261.5	1289348.6	2049478.5
22	0.99	13405.25	1	1	1	59498.6	17851.9	5632728.2	4130768.6
23	0	0	1	1	1	0	0	0	0
24	2.67	32321.56	3.89	0.06	2.63	164691.3	48254.2	24789869.3	7970440.0
25	4.19	0	1	0.01	3.42	135954.2	87756.1	78591479.3	10538883.0

26	6.98	0	1	1	1	667062.8	102649.9	77879216.7	0
27	0.01	12.38	1.55	0.52	1.31	0	0	8419.8	120856.8
28	0	0	1	1	1	0	0	0	0
29	0.90	1175.51	2.53	0.13	1.81	25284.3	3945.7	5358054.1	8512553.0
30	0.80	4942.57	3.51	0.04	2.94	13201.4	11711.7	6091569.7	6580415.2
31	0	0	1	1	1	0	0	0	0
32	1.06	0	1	0.13	1.61	71365.6	19231.4	15664061.6	1065587.7
33	0.84	0	1	0.30	1	41410.5	21791.7	15997357.9	0
34	0.79	0	1	0.07	2.38	32795.1	19401.7	11810380.9	1603358.0
35	1.58	0	1	0.04	1.79	94286.7	31418.5	20441613.7	2861354.5
36	2.64	0	1	0.01	3.41	154671.9	38838.6	33608001.5	7673041.9
37	0.95	0	1	1	1	2021.7	8692.7	1778084.5	11489805.0
38	1.70	0	1	0.17	1	102754.9	39950.3	29074074.2	0
39	0.24	87.44	5.45	0.02	4.74	7520.4	0	988658.4	2448874.8
40	0.11	2064.94	1	1	1	2071.9	1596.7	0	1061280.0
41	1.12	15285.46	1	0.06	1.62	42292.5	27202.9	10253152.3	4654606.5
42	0.97	12193.91	1	0.51	1.19	72303.1	24566.3	11002835.6	0
43	1.06	6881.97	1	0.05	3.13	10592.0	13899.7	4174428.4	10720333.3
44	2.00	0	1	0.02	3.57	149199.7	28691.9	13270919.8	6407825.6
45	0.96	0	1	0.07	3.41	59.2	1651.3	2771503.4	12739306.0
46	0.21	0	1	1	1	0	1218.6	125186.5	2771301.6
47	1.34	566.78	3.68	0.02	3.68	59555.0	3614.0	2703594.9	12472014.0
48	0	0	1	1	1	0	0	0	0
49	3.24	0	1	0.00*	9.25	152817.4	73929.2	45012346.0	7089624.9
50	1.72	0	1	1	1	62564.8	69649.1	23477832.3	245932.0
51	0.71	4760.41	14.36	0.03	5.28	5964.9	3658.1	0	8986034.4
52	4.70	0	1	0.00*	5.48	307320.3	74482.9	54969087.3	11266074.0

(continued)

Table 11.4 (continued)

DMU #	MAM	M1 (DDF)	M2	M3	M4 (HEM)	Optimal slacks of the additive model (MAM)			
						SO_2	NO_X	CO_2	Total sales
53	0	0	1	1	1	0	0	0	0
54	0	0	1	1	1	0	0	0	0
55	0	0	1	1	1	0	0	0	0
56	1.22	9623.65	5.18	0.02	4.16	29703.7	22451.8	12783408.3	7339858.7
57	0.06	162.15	3.86	0.24	2.04	0	57.07	0	918601.5
58	0.49	9575.23	1	1	1	16528.5	16197.7	5229772.2	1248305.9
59	0.02	0	1	1	1	0	1566.29	0	48377.3
60	4.02	0	1	0.06	1	121810.1	134441.9	81747470.3	176624.0
61	1.23	0	1	0.05	1.29	59555.0	3614.0	2703594.9	10881532.0
62	1.00	0	1	0.00*	185.20	53478.9	16585.9	12289400.2	3228289.1
63	0.40	3615.73	1.13	0.45	1.13	13862.8	10253.5	5962576.3	958514.0
64	2.27	0	1	1	1	245600.4	28771.6	18758545.0	0
65	1.65	4105.45	1.29	0.12	1.28	50063.6	48366.1	29611177.3	2377461.0
66	0.84	0	1	0.02	2.12	46276.0	1197.4	3677057.7	6692419.6
67	0.54	4984.54	1	1	1	7705.3	11753.8	3453290.2	4061772.0
68	1.42	14765.38	6.99	0.01	5.08	39505.5	24479.9	12073234.4	9195619.6
69	1.20	0	1	0.06	3.14	12068.8	19493.8	11702312.3	9470109.0
70	0.37	0	1	1	1	21345.0	529.5	0	3305391.2
71	0	0	1	1	1	0	0	0	0
72	0.84	4409.70	16.17	0.02	6.89	5673.1	6982.0	2280865.7	9632715.3
73	1.43	10437.80	3.92	0.02	2.78	94048.5	17678.0	11728859.2	5634588.3
74	0.62	0	1	1	1	4085.8	14664.7	3180300.0	5017813.0
75	0.99	0	1	1	1	87828.1	12193.6	10070147.6	1249568.4
76	1.05	0	1	1	1	40940.7	16452.5	19883465.5	3000247.9

77	0.94	0	1	0.20	1	37227.1	17623.3	16084807.6	2594252.7
78	1.60	0	1	0.01	6.23	19440.8	48750.9	24833943.2	5711590.2
79	0.55	1027.08	3.05	0.14	2.26	0	1869.5	948663.0	7268626.1
80	0.42	4025.90	1.28	0.30	1.35	9556.9	12731.2	5643480.8	1357680.7
81	0	0	1	1	1	0	0	0	0
82	0.35	174.04	8.73	0.06	4.01	516.2	0	237477.1	4878351.2
83	3.51	27724.28	3.97	0.01	3.71	162057.3	74328.3	49044620.3	8911863.0
84	0.47	597.88	3.47	0.13	2.52	913.0	2097.3	1203764.1	5906266.5
85	0	0	1	1	1	0	0	0	0
86	0	0	1	1	1	0	0	0	0
87	1.60	1274.70	1.80	0.02	2.56	54143.4	22443.5	24121696.2	7414865.6
88	1.64	3906.81	2.98	0.02	2.34	94729.5	15905.1	14490390.8	8097179.2
89	0.99	7322.13	3.68	0.04	2.69	28897.6	10300.7	5717873.3	8127610.4
90	0.97	348.33	1.03	0.34	1.03	57669.9	15743.2	16796866.3	1262630.0
91	0	0	1	1	1	0	0	0	0
92	0	0	1	1	1	0	0	0	0
93	1.60	17197.03	4.83	0.02	3.56	64718.2	26006.2	12952917.9	8980738.4
94	0	0	1	1	1	0	0	0	0

The number with mark '*' is very small but still strictly larger than zero

Table 11.5 Efficiency results of projection points

DMU #	M1 (DDF)	M2	M3	M4 (HEM)
1	3.68	3.13	1.68	4.41
2	1.17	1.07	2.92	1.68
3	0.75	0.63	0.10	0.20
4	0.54	0.12	0.28	0.23
5	0.09	0.09	0.22	0.13
6	0	0	0	0
7	0.66	0.55	0.00*	0.00*
8	0.76	0.11	0.01	0.05
9	0	0	0	0
10	0	0	0	0
11	2.17	1.08	1.23	0.68
12	2.21	1.73	0.04	0.62
13	1.57	1.22	2.25	1.65
14	0.30	0.06	0.00*	0
15	2.86	1.63	0.00*	0.45
16	1.78	0.21	0.93	1.14
17	0.43	0.36	0.60	0.36
18	6.17	5.65	0.39	0.43
19	2.61	2.46	7.91	4.11
20	2.41	2.16	0.01	0.17
21	0.22	0.18	0.32	0.21
22	0.77	0.94	2.65	1.40
23	0	0	0	0
24	2.16	0.43	0.39	0.90
25	4.46	4.05	0.00*	1.56
26	7.51	6.98	22.83	11.57
27	0.00	0.00*	0	0.00*
28	0	0	0	0
29	0.90	0.33	0.12	0.38
30	0.73	0.31	0.04	0.14
31	0	0	0	0
32	1.14	1.05	0.26	0.93
33	0.91	0.84	0.23	1.33
34	0.85	0.76	0.05	0.32
35	1.69	1.54	0.12	0.84
36	2.80	2.54	0.01	0.67
37	0.97	0.80	1.20	0.79
38	1.84	1.70	0.39	2.73
39	0.24	0.13	0.01	0.04
40	0.06	0.09	0.22	0.12
41	0.88	1.06	0.17	0.84
42	0.80	0.97	1.13	0.94

(continued)

Table 11.5 (continued)

DMU #	M1 (DDF)	M2	M3	M4 (HEM)
43	0.95	0.92	0.11	0.35
44	2.13	1.91	0.23	0.77
45	0.96	0.79	0.00	0.05
46	0.21	0.17	0.24	0.16
47	1.36	0.96	0.06	0.32
48	0	0	0	0
49	3.48	3.14	0.02	0.21
50	1.92	1.72	4.72	2.63
51	0.50	0.47	0.06	0.20
52	5.01	4.55	0.03	0.90
53	0	0	0	0
54	0	0	0	0
55	0	0	0	0
56	1.08	0.56	0.02	0.16
57	0	0.01	0.00*	0.00*
58	0.33	0.48	1.22	0.69
59	0.03	0.02	0.05	0.03
60	4.39	4.01	0.33	6.23
61	1.26	1.09	0.16	1.06
62	1.07	0.96	0.00*	0.01
63	0.35	0.26	0.28	0.32
64	2.44	2.27	7.70	3.79
65	1.70	1.40	0.30	1.62
66	0.87	0.75	0.06	0.53
67	0.47	0.49	0.98	0.59
68	1.18	0.73	0.05	0.22
69	1.25	1.08	0.07	0.11
70	0.38	0.33	0.77	0.41
71	0	0	0	0
72	0.76	0.23	0.02	0.10
73	1.29	0.76	0.11	0.54
74	0.65	0.55	1.04	0.64
75	1.05	0.97	3.14	1.60
76	1.11	1.01	2.63	1.53
77	1.00	0.91	0.31	1.35
78	1.73	1.52	0.04	0.20
79	0.52	0.05	0.03	0.05
80	0.37	0.32	0.23	0.28
81	0	0	0	0
82	0.33	0.07	0.00*	0
83	3.16	2.34	0.02	0.38
84	0.46	0.02	0.03	0.04

(continued)

Table 11.5 (continued)

DMU #	M1 (DDF)	M2	M3	M4 (HEM)
85	0	0	0	0
86	0	0	0	0
87	1.65	1.29	0	0.57
88	1.63	0.98	0.03	0.75
89	0.86	0.14	0.06	0.22
90	1.03	0.93	0.76	1.42
91	0	0	0	0
92	0	0	0	0
93	1.32	0.64	0.11	0.38
94	0	0	0	0

The number with mark '*' is very small but still strictly larger than zero

11.5 Conclusions

In this chapter, we examine three critical implementation issues of the DDF and radial efficiency models which are widely used for the environmental efficiency evaluation in the literature: *non-monotonicity*, *misclassification of efficiency status*, and *strongly dominated projection targets*. Our analysis shows that the classical weak disposability assumption on undesirable outputs can create a portion of the output-dominated frontier, which can be considered the root cause for the three issues. Our findings provide important implications for both empirical and theoretic researchers of environmental efficiency. We suggest that researchers should be cautious when imposing the classical weakly disposability assumption on undesirable outputs under both CRS and VRS production technologies, which has been the standard assumption in a large stream of studies.

As the importance of environmental efficiency is growing, findings from this study have an important theoretical implication. Further, the application areas of the environmental efficiency model can be applied to many other dimensions of corporate operations when both positive and negative consequences of an activity or policy (e.g., debts in banking, labor accidents and litigations in transportation and manufacturing). Researchers are encouraged to explore more application areas in other emerging contexts.

References

Banker RD, Charnes A, Cooper WW (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis. Manag Sci 30(9):1078–1092

Baucells M, Sarin RK (2003) Group decisions with multiple criteria. Manag Sci 49(8):1105–1118

Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision making units. Eur J Oper Res 2(6):429–444

Charnes A, Cooper WW, Golany B, Seiford L, Stutz J (1985) Foundations of data envelopment analysis for Pareto-Koopmans efficient empirical production functions. J Econometrics 30 (1):91–107
Chen CM (2013) A critique of non-parametric efficiency analysis in energy economics studies. Energy Econ 38:146–152
Chen CM (2014) Evaluating eco-efficiency with data envelopment analysis: an analytical reexamination. Ann Oper Res 214(1):49–71
Chen CM, Delmas M (2011) Measuring corporate social performance: an efficiency perspective. Prod Oper Manag 20(6):789–804
Chen CM, Delmas MA (2012) Measuring eco-inefficiency: a new frontier approach. Oper Res 60 (5):1064–1079
Chung YH, Färe R, Grosskopf S (1997) Productivity and undesirable outputs: a directional distance function approach. J Environ Manage 51(3):229–240
Cooper WW, Park KS, Pastor JT (1999) RAM: a range adjusted measure of inefficiency for use with additive models, and relations to other models and measures in DEA. J Prod Anal 11 (1):5–42
Cooper WW, Seiford LM, Tone K (2007) Data envelopment analysis: a comprehensive text with models, applications, references and DEA-solver software. Springer, New York
Cooper WW, Pastor JT, Borras F, Aparicio J, Pastor D (2011) BAM: a bounded adjusted measure of efficiency for use with bounded additive models. J Prod Anal 35(2):85–94
Delmas M, Russo MV, Montes-Sancho MJ (2007) Deregulation and environmental differentiation in the electric utility industry. Strateg Manag J 28(2):189–209
Fabrizio KR, Rose NL, Wolfram CD (2007) Do markets reduce costs? Assessing the impact of regulatory restructuring on US electric generation efficiency. Am Econ Rev 97(4):1250–1277
Färe R, Grosskopf S (2003) Nonparametric productivity analysis with undesirable outputs: comment. Am Econ Rev 85(4):1070–1074
Färe R, Grosskopf S (2006) New directions: efficiency and productivity. Springer, New York
Färe R, Grosskopf S (2010) Directional distance functions and slacks-based measures of efficiency. Eur J Oper Res 200(1):320–322
Färe R, Grosskopf S, Lovell CK, Pasurka C (1989) Multilateral productivity comparisons when some outputs are undesirable: a nonparametric approach. Rev Econ Stat 71(1):90–98
Farrell MJ (1957) The measurement of productive efficiency. J Royal Stat Soc Ser A (General) 120 (3):253–290
Kuosmanen T (2005) Weak disposability in nonparametric production analysis with undesirable outputs. Am J Agric Econ 87(4):1077–1082
Kuosmanen T, Podinovski V (2009) Weak disposability in nonparametric production analysis: reply to Färe and Grosskopf. Am J Agric Econ 91(2):539–545
Majumdar SK, Marcus AA (2001) Rules versus discretion: the productivity consequences of flexible regulation. Acad Manage J 44(1):170–179
Mandal SK, Madheswaran S (2010) Environmental efficiency of the Indian cement industry: an interstate analysis. Energy Policy 38(2):1108–1118
Oggioni G, Riccardi R, Toninelli R (2011) Eco-efficiency of the world cement industry: a data envelopment analysis. Energy Policy 39(5):2842–2854
Riccardi R, Oggioni G, Toninelli R (2012) Efficiency analysis of world cement industry in presence of undesirable output: application of data envelopment analysis and directional distance function. Energy Policy 44:140–152
Seiford LM, Zhu J (2002) Modeling undesirable factors in efficiency evaluation. Eur J Oper Res 142(1):16–20
Seiford LM, Zhu J (2005) A response to comments on modeling undesirable factors in efficiency evaluation. Eur J Oper Res 161(2):579–581

Shephard RW (1970) Theory of cost and production functions. Princeton University Press, Princeton

Steinmann L, Zweifel P (2001) The range adjusted measure (RAM) in DEA: comment. J Prod Anal 15(2):139–144

Zhou P, Poh KL, Ang BW (2007) A non-radial DEA approach to measuring environmental performance. Eur J Oper Res 178(1):1–9

Chapter 12
Applications of Data Envelopment Analysis in Education

Emmanuel Thanassoulis, Kristof De Witte, Jill Johnes, Geraint Johnes, Giannis Karagiannis, and Conceição S. Portela

Abstract Non-parametric methods for efficiency evaluation were designed to analyse industries comprising multi-input multi-output producers and lacking data on market prices. Education is a typical example. In this chapter, we review applications of DEA in secondary and tertiary education, focusing on the opportunities that this offers for benchmarking at institutional level. At secondary level, we investigate also the disaggregation of efficiency measures into pupil-level and school-level effects. For higher education, while many analyses concern overall

---

Part of this research has been co-financed by the European Union (European Social Fund ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) – Research Funding Program: THALES. Investing in Knowledge Society through the European Social Fund. The views expressed in this chapter are those of the authors and no representation is made that they are shared by any funding body.

E. Thanassoulis (✉)
Aston Business School, Aston University, Birmingham, UK
e-mail: e.thanassoulis@aston.ac.uk

K. De Witte
Leuven Economics of Education Research, Faculty of Economics and Business,
KU Leuven, Leuven, Belgium

Top Institute for Evidence Based Education Research, Maastricht University,
Maastricht, The Netherlands

J. Johnes
Huddersfield University Business School, University of Huddersfield, Huddersfield, UK

G. Johnes
Lancaster University Management School, Lancaster University, Lancaster, UK

G. Karagiannis
Department of Economics, University of Macedonia, Thessaloniki, Greece

C.S. Portela
Católica Porto Business School, Porto, Portugal

J. Zhu (ed.), *Data Envelopment Analysis*, International Series in Operations
Research & Management Science 238, DOI 10.1007/978-1-4899-7684-0_12

institutional efficiency, we examine also studies that take a more disaggregated approach, centred either around the performance of specific functional areas or that of individual employees.

Keywords DEA • Efficiency • Education • Benchmarking • Pupil-level effects • School efficiency • Higher education efficiency

12.1 Introduction

Data envelopment analysis (DEA) was originally developed to provide a means of efficiency evaluation in the context of 'not-for-profit entities participating in public programs' (Charnes et al. 1978). Not all such entities are providers of public goods – in the sense that their output is non-rival and non-excludable – but they are all characterised by a production process that converts a multiplicity of inputs into a multiplicity of outputs for which market prices are absent. In this respect, education represents a classic example of a sector which is well served by DEA.

Beyond elementary education, schools, colleges and universities provide specialist tuition in a wide variety of subject areas. Inputs include the expertise of teachers in each of these subjects, and the extent of their specialism means that teaching in each subject (or at least in each cluster of subjects) should be regarded as a distinct input. Likewise, a multiplicity of outputs reflects students who (at secondary level) might take a vocational or academic route, or who (at tertiary level) specialise in particular disciplines. For many higher education institutions, research provides a further distinct output. There exists considerable synergy between the various activities, and cross-subsidisation is common.

The education sector in most countries comprises both public and private provision. In the private sector as much as in the public, joint production of multiple outputs is common. Given unobserved heterogeneity across students, assessing the benefit of education for any individual student is problematic, and in consequence the price charged for educational services does not have many of the characteristics that are associated with market pricing. The presence of multiple inputs, multiple outputs, and prices that are unlikely to serve a useful purpose as weights, are features that combine to make DEA an instructive tool in this context.

The structure of this chapter is as follows. In the next section, we look at applications of DEA in secondary education. These include studies that operate at various levels of aggregation. First we review studies that focus on relatively highly aggregated data, evaluating the efficiency of schools. Typically these include as inputs the characteristics of schools and their pupil intakes; meanwhile outputs include various measures of pupil attainment. A particularly interesting development in this area is the emergence of online platforms that allow schools to enter data about their own performance and then compare this performance with that of peers (who have likewise entered data onto the platform). This illustrates very vividly the scope for DEA and similar methodologies to provide benchmarking information that can be useful in the dissemination of good practice and hence in the process of securing efficiency improvements. We next proceed to review studies that focus on the pupil as the level of analysis. These are relatively rare, partly because of data availability issues, but partly also because the computational burden of DEA becomes considerable when dealing with large numbers of individual pupils. However these studies are insightful, not least because they allow each school's frontier to be understood as an envelope around its students' performance. Our assessment of the relative performance of two otherwise identical students attending different schools might be conditioned by information about differences in the frontiers associated with the schools that they attend. This separation into a school effect and a pupil effect has some aspects in common with the statistical approach of multilevel modelling, and we draw comparisons.

In reviewing the literature on school efficiency, we consider also the way in which efficiency has been observed to change over time, and in particular focus on the extent to which such change is due to either change in the distribution of efficiencies or movements of the efficiency frontier itself.

In Sect. 12.3, we review studies of DEA as applied to higher education institutions, focusing on analyses that use data aggregated to the level of the institution. These studies include analyses of the cost efficiency and technical efficiency of overall operations. We also consider studies that have focused on particular aspects of university activities, specifically including investigation of the efficiency of administrative services and the efficiency of research production in the university sector.

The basic DEA model has been extended to study a variety of applications in higher education, and some of these are reviewed in this section of the chapter. For instance, there are some examples of merger activity in higher education – (concern about) efficiency is often cited as a cause and (a change in) efficiency is an effect of this activity. Higher education is also one of the areas in which network DEA has been applied, yielding results that are informative about the sources of relative inefficiency within the 'black box' of production.

More disaggregated data have been used in the higher education sector to evaluate the performance of individual staff members along a number of dimensions. We review these studies in Sect. 12.4, focusing specifically first on academics' research output, and second on their teaching output.

The chapter ends with a conclusion that draws together the main threads of the preceding discussion.

12.2 Applications of DEA in Secondary Education

12.2.1 Introduction

Educational data follow typically a hierarchical structure, since pupils are nested within classes, classes are nested within schools, schools are nested within districts, school districts are nested within states, which are nested within countries. Within each of these levels there are variables of interest for educational policies (e.g. pupils' socio-economic background, ability of peers in the same class, size of the school, state policies regarding the autonomy of schools etc.). The analysis of data at several levels requires the adaptation of existing parametric or non-parametric techniques which are usually designed for single level analysis.

Since pupil-level data are at the lowest disaggregated level, the analysis of these data can provide more information to practitioners and researchers. As a result, most school effectiveness research is actually undertaken at this level of analysis. Research on school effectiveness started in the 60s. Its most influential article is the controversial Coleman Report (1966). Conclusions of this report pointed to the lack of importance of the schools themselves in explaining attainment by pupils. This, being a counter intuitive finding, gave rise to a number of studies whose aim was to prove that schools did make a difference. These studies typically approach education as a production process, where student outcomes are a function of several variables. The 'educational production function' framework groups inputs into four main dimensions: Family background, Peer influences; School inputs; and Innate abilities of students (see Hanushek 1979). The variables considered within this general model are constructs that need to be operationalized using specific quantitative measures.[1] For example, for operationalizing the innate abilities prior attainment before entering a given stage of education under assessment is usually used as a proxy. The use of prior attainment to explain subsequent attainment gave rise to what have been termed value-added (**VA**) studies. Such studies differ from other school effectiveness studies as they focus on the progress schools help pupils to make relative to their different starting points (see e.g. Meyer 1997; Goldstein et al. 2000). As a result, VA studies have a longitudinal perspective of pupil attainment: they are usually undertaken at the pupil-level and consider the achievements of pupils over a certain cycle of studies, with achievements on entry and on exit of that cycle as the main variables of interest. For detailed discussions around VA models and implementations see OECD (2008). Clearly several other variables can be considered within the analysis. Typically socio-economic characteristics of the students are considered as an additional and very important driver of exit achievement. Note however, that consideration of prior (entry) achievement "implicitly controls for socio-economic status and other background factors to the extent that their influence on the post-test is already reflected in the pre-test score" (Ballou et al. 2004, p. 38).

Assessments of the comparative efficiency of secondary schools in value-added fall into two broad categories in terms of the data used. Those using aggregate and those using pupil-level data. The type of data used is a combination of data availability and the issues to be addressed through the assessment. We outline sample applications from both genres. However, in the context of DEA, earlier

[1] Clearly qualitative studies are of utmost importance in education, but we address here only quantitative studies.

applications used aggregate pupil data (i.e. school data or school districts data) and so we look first at DEA applications using these types of data. Note that within this type of studies a schools' VA perspective can be taken, but an efficiency perspective may also be investigated. Banker et al. (2004) distinguished between the assessment of efficiency and effectiveness in schools, which could be operationalized by choosing different types of variables for the assessment. Mayston (2003) considers a similar classification of school studies, where the distinction lies in the consideration or not of expenditures and school resources. In this chapter we take efficiency to be 'value for money' (i.e. where school resources including expenditures are central), while effectiveness is value-added (i.e. where school resources and other 'endogenous' variables are not permitted to explain exit attainment). In this respect we adopt the Mayston (2003) distinction between efficiency and effectiveness in education assessments.

12.2.2 Applications Using Aggregate Pupil Data

Aggregate pupil data are more readily amenable to DEA than pupil-level data and this explains in large measure why DEA assessments initially relied on such data. Pupil-level data, as we will see later, offer greater scope for addressing questions such as identifying pupil-level as distinct from school-level effects on VA. Nevertheless, the readily available aggregate data also make it possible to address significant questions about school effectiveness and efficiency as we now illustrate.

12.2.2.1 An Overview of DEA Applied to Aggregate Pupil Data

The number of DEA applications on aggregate pupil data has grown since the 80s when the first studies of this type emerged. The first school efficiency study was that of Bessent and Bessent (1980), followed by a very influential paper of Charnes et al. (1981) that will be covered in some detail below. An extensive survey of the earlier literature is provided by De Witte and López-Torres (2015).

The DEA studies that have used the school as the level of analysis have used in general standard DEA models (except in a few cases), differing mainly on the type of inputs and outputs used, and therefore on the focus of the analysis. A general consensus regarding the type of inputs that should be considered in such studies has emerged in the literature as three groups of variables are usually considered: (i) those reflecting characteristics of pupils (like prior attainment and socio-economic characteristics), (ii) those reflecting characteristics of the school (like number of teaching and non-teaching staff, expenditure per pupil, size of school, or class size), and (iii) those reflecting characteristics of teachers (like their salary, experience, or level of education). Regarding outputs, the consensus is that these should relate to standardized test scores, but they have been aggregated in several forms like the median (Bessent and Bessent 1980), the mean (Mizala et al. 2002;

Table 12.1 Input/output set in Charnes et al. (1981)

Inputs	Outputs
Maternal education level	Attainment in reading
Highest occupation of a family member	Attainment in maths
Parental visits to school	Self esteem
Number of teachers	

Muñiz 2002), or the proportion of pupils achieving more than a certain grade (Bradley et al. 2001). Other relevant outputs also related to pupils' achievement are the number of approvals or success rates (Kirjavainen and Loikkanen 1998; Muñiz 2002; Oliveira and Santos 2005), attendance rate (Arnold et al. 1996; Bradley et al. 2001), number of graduates (Kirjavainen and Loikkanen 1998), and percentage of students who do not drop out from school (Arnold et al. 1996).

One of the studies that can be classified within a school effectiveness perspective that gave rise to subsequent applications of DEA in education (both at the school and at the pupil-level) is that of Charnes et al. (1981). This application uses DEA in the context of what could be called a 'matched-school' experiment and it has led to what became known as disentangling 'managerial' from 'program' efficiency. The approach spawned a plethora of applications both within and outside education.

The study of Charnes et al. (1981) originated in the USA, where federal authorities at that time wished to test the effectiveness of a program of interventions with primary school children known as 'Program Follow Through' (PFT). The aim of PFT was to compensate disadvantaged children by instituting academic and indeed non-academic interventions (e.g. social, nutritional and other counselling), which would be offered to 'treated' cohorts of children. Each 'treated cohort' was matched with an 'untreated' cohort. The intention of the study was to establish whether PFT was achieving its aim of compensating to some degree for the disadvantaged background of the children concerned.

The study used data both from treated and untreated cohorts, normalised as per 100 pupils. The input/output set used in that study is shown in Table 12.1, where suitably constructed measures for the cohort were chosen.

The key aim of the study was to isolate through DEA any ineffectiveness in implementation of the PFT so that the effectiveness of the program itself could be identified. The approach used is illustrated graphically below.

Let two sets of units (cohorts of pupils in the case of the study under consideration) operate under 'program' 1 and 2 respectively. Assume the units operate under constant returns to scale, using two inputs to secure one output. In Fig. 12.1 ACD is the efficient boundary of program 1 and EFG that of program 2. The efficiency of each unit relative to its own policy boundary is its 'managerial' efficiency.

Thus OB/OJ is the managerial efficiency of school J of Program 1 and OK′/OK that of school K operating within Program 2. To discount managerial inefficiencies and isolate 'program' efficiencies each school is projected to the efficient boundary of its policy as depicted in Fig. 12.2.

EFCD is a global or meta frontier enveloping all programs. OM″/OM′ is the 'program 1' efficiency at the input mix of unit M. The overall inefficiency of unit M

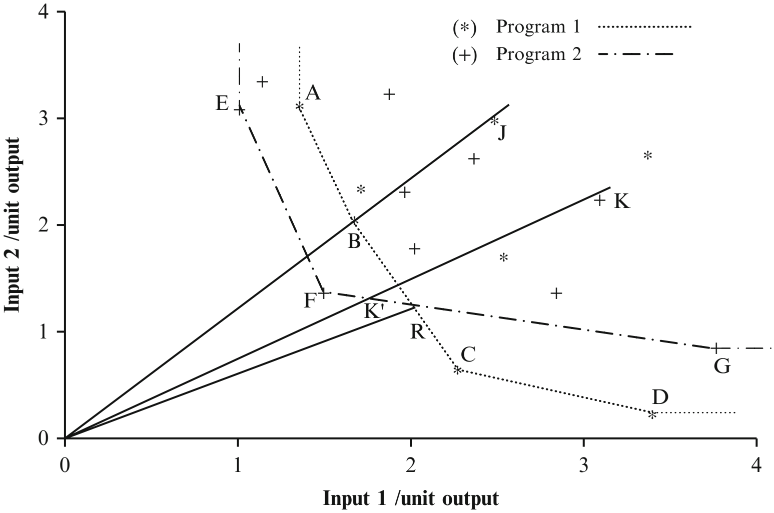


Fig. 12.1 Efficient boundaries drawn by policy

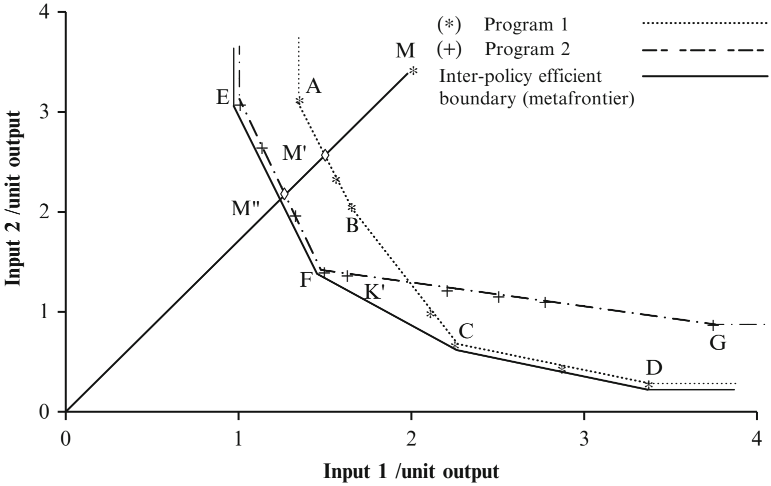


Fig. 12.2 Global (meta) frontier envelops the program boundaries

is OM″/OM. This is decomposed into OM″/OM′ attributable to the program under which unit M operates (Program 1 in this case) and OM′/OM is the component of the inefficiency of Unit M that is attributable to its own management. (Note that though here the overall inefficiency of unit M, OM″/OM is multiplicatively decomposed so that OM″/OM = (OM″/OM′) × (OM′/OM) this will not generally be the case where the efficient projection of a unit such as M is such that there are slacks to be eliminated in order to render that projection Pareto efficient).

The process illustrated above can be operationalised using multiple inputs and outputs by solving the DEA models (12.1) and (12.3) as explained below.

Assume that there are N_p DMUs operating under program p ($p = 1...P$), and that DMU j of program p uses inputs x_{ij}^{p} (i = 1,...,m) to secure outputs y_{rj}^{p} (r = 1,...,s). The managerial technical input efficiency of DMU j_0 of program p is the optimal value of k_{j0}^{p} in (12.1).

$$
\begin{aligned}
& Min\, k_{j0}^{p} - \varepsilon\left(\sum\nolimits_i S_i^- + \sum\nolimits_r S_r^+\right) \\
& st: \\
& \sum_{j=1}^{n} \lambda_j x_{ij}^{p} = k_{j0}^{p} x_{ij0}^{p} - S_i^- \quad i = 1, \ldots, m \\
& \sum_{j=1}^{n} \lambda_j y_{rj}^{p} = y_{rj0}^{p} + S_r^+ \qquad r = 1, \ldots, s \\
& \lambda_j \geq 0, \forall j, \quad S_i^-, S_r^+ \geq 0, \quad \forall i, \text{ and } r, k_{j0}^{p} \text{ free}
\end{aligned}
\tag{12.1}
$$

Let ($x_{ij_0}^{tp}$, $i = 1\ldots m$, $y_{rj_0}^{tp}$, $r = 1\ldots s$) be a set of input-output levels that would render DMU j_0 Pareto-efficient within its program p. We will use the set ($x_{ij_0}^{tp}$, $i = 1\ldots m$, $y_{rj_0}^{tp}$, $r = 1\ldots s$) yielded by model (12.1) so that:

$$
\begin{aligned}
x_{ijo}^{tp} &= \sum_{j=1}^{n} \lambda_j^* x_{ij}^{p} = k_{jo}^{p*} x_{ijo}^{p} - S_i^{-*} \quad i = 1, \ldots, m \\
y_{rjo}^{tp} &= \sum_{j=1}^{n} \lambda_j^* y_{rj}^{p} = y_{rjo}^{p} + S_r^{+*} \qquad r = 1, \ldots, s
\end{aligned}
\tag{12.2}
$$

where the superscript * denotes the optimal value of the corresponding variable in (12.1).

The program efficiency at the input mix of DMU j_0 is the optimal value p_0^* of p_0 in model (12.3).

$$
\begin{aligned}
& Min \qquad p_0 - \varepsilon\left(\sum\nolimits_i S_i^- + \sum\nolimits_r S_r^+\right) \\
& st: \\
& \sum_{p=1}^{P}\sum_{j=1}^{N_p} \lambda_j x_{ij}^{tp} = p_0 x_{ij0}^{tp} - S_i^- \quad i = 1, \ldots, m \\
& \sum_{p=1}^{P}\sum_{j=1}^{N_p} \lambda_j y_{ij}^{tp} = y_{rj0}^{tp} + S_r^+ \qquad r = 1, \ldots, s \\
& \lambda_j \geq 0, \forall j, \quad S_i^-, S_r^+ \geq 0, \quad \forall i, \text{ and } r, \; p_0 \text{ free}
\end{aligned}
\tag{12.3}
$$

Charnes et al. (1981) used the above approach for illustrative rather than definitive purposes in order to disentangle managerial from program efficiencies in the case of PFT and untreated cohorts labelled NFT (non follow through). They point in particular how 'inter-program' areas such as the solid boundary FC in Fig. 12.2 can indicate other amalgams of Programs not constituted initially that can give useful indications of interventions that might be constructed as new Programs.

To the extent that all schools have as basic aim enhancing the educational attainments of pupils but at the same time pupils and/or schools often operate in different contexts (e.g. they may operate under different regimes (e.g. fee charging vs publicly funded schools)) the approach outlined above isolates the impact of context from the effectiveness of the school itself, when we control for context. The approach has been adapted in many educational contexts, e.g. see Portela and Thanassoulis (2001); Thanassoulis and Portela (2002); De Witte et al. (2010); Mancebón et al. (2012). Further, the same approach has been used in many other areas away from education such as in banking (e.g. Golany and Storbeck 1999; Johnes et al. 2014), and water (e.g. De Witte and Marques 2009).

Some studies measuring school effectiveness did not consider the approach of Charnes et al. (1981) but applied in general a non-parametric methodology to the estimation of school effectiveness. An example can be found in Cherchye et al. (2010). These authors used as inputs the total number of instruction units assigned to a particular pupil. For their Flemish application, this consists of regular (REG) and additional, so-called 'equal educational opportunity' (EEO), instruction units (depending on certain 'disadvantageous' pupil characteristics). Output is defined on the basis of test scores in three dimensions: mathematics, technical reading and writing, collected at the end of the second year.

Regarding school efficiency studies these are distinguished from effectiveness studies by the fact that a central consideration is that of school expenditures on the input side of the assessment. The aim is to assess the extent to which some schools or school districts are more cost effective than others in providing school outcomes. On the input side of such assessments, the inputs relating to prior attainment or to the socio-economic characteristics of pupils may also be considered as they too impact pupil outcomes.

An example of an early study on school efficiency is that by Färe et al. (1989), where the authors considered 40 Missouri school districts, using on the input side variables such as the number of students, the net expenditure and the number of eighth grade teachers while outputs were the number of students passing three types of exams. Another study of the same type is that of Ruggiero (1999) where the author addressed explicitly the cost efficiency of 584 school districts in New York. Conclusions point to the striking figure of 64 % of school districts being cost inefficient, (in Färe et al. (1989) more than half of the school districts were considered efficient). Ruggiero (1999) used a single input in a DEA model (the expenditure per pupil) and considered as environmental variables the percentage of minority students and the percentage of limited English students (a measure of the effect of the environment on costs was estimated in this study). Fukuyama and Weber (2002) also analysed cost efficiency (and allocative efficiency) of 310 Texas school districts using several model specifications. They divided inputs into variable (number of administrators, teachers, teacher aides and support staff) and fixed (predicted achievement of pupils based on their prior achievement and socio-economic characteristics and operating expenses per pupil). Outputs used were the value added test scores (see Grosskopf et al. 1997 on how these measures are obtained). The approach of Fukuyama and Weber is an extension of the previous work of Grosskopf et al. (1999) where the authors used an indirect distance function that allowed the measurement of output expansion possible if school districts were able to re-allocate inputs while maintaining a given budget. Banker et al. (2004) also examined efficiency of Texas school districts. They used as inputs three types of expenditures (related to: instruction; administration; and support), and used as outputs the total pupil enrolment in elementary schools, middle schools and high schools.

The foregoing examples relate to the US, where assessments are normally at district level. Elsewhere studies at school level can be found for example in Burney et al. (2013) who look at the efficiency of public schools in Kuwait. Haelermans and De Witte (2012) examined the influence of innovations in the efficiency of Dutch secondary schools. They observed that profiling, pedagogic, process and education chain innovations are significantly related to school efficiency, whereas innovations in the professionalization of teachers are insignificantly related to school efficiency. Portela et al. (2012) also looked at efficiencies of Portuguese schools. We return to this assessment in Sect. 12.2.2.3 as it is linked with an online tool that allows the computation of school efficiency scores online. Another study in Portuguese schools is that in Portela and Camanho (2007) which is a good example of how the two perspectives of assessment, efficiency and effectiveness can be seen as complementary. The authors assessed schools from two perspectives: One called the society perspective (related to effectiveness studies mentioned above), and the other called the educational authorities' perspective (related to efficiency studies mentioned above). If one is evaluating schools from the parents' perspective, the fact that some schools may appear to have low Value Added due to scarce resources, poor location, or poor quality of teachers, is of no particular importance. The objective of parents is just to identify the best schools for their children rather than make allowances for less than satisfactory performance by teachers or schools. However, if one is assessing the schools from the perspective of an authority charged with funding and overseeing the services delivered by schools (the most usual implicit perspective in published papers) all the reasons behind poor or excellent VAs are of interest if steps are to be taken to improve the performance of the schools. In this case an efficiency perspective should be also considered.

12.2.2.2 Using Aggregate Pupil Data to Identify Differential School Effectiveness

It has been found (e.g. Gray et al. 1986; Sammons et al. 1993) that some schools have differential effectiveness depending on pupil prior attainment levels. More generally a school may intentionally or otherwise be using teaching styles which favour more certain groups of pupils compared to others.

Identification of the direction and degree of any differential effectiveness at schools is valuable for a number of reasons. Thanassoulis (1996) argues it helps to identify suitable teaching practices and role model schools for improved all round effectiveness. Where a school streams pupils by ability, appropriate teaching practices and role model schools for raising the level of achievement of pupils of a given ability range can be identified. See also Sammons et al. (1993) on the implications of differential effectiveness for comparing schools. It is preferable to use pupil-level data to ascertain the existence or otherwise of differential school effectiveness as illustrated in the next section.

Using pupil-data, De Witte and Van Klaveren (2014) examine which configuration of teaching activities maximizes student performance. Using data from the Trends in International Mathematics and Science Study, they formulate a nonparametric efficiency model that accounts for self-selection of students and teachers in better schools, and complementary teaching activities. The analysis distinguishes both individual teaching (i.e. a personal teaching style adapted to the individual needs of the student) and collective teaching (i.e. a similar style for all students in a class). Moreover, they test to which group of students the teacher is adapting his/her teaching style. De Witte and Van Klaveren (2014) show that high test scores are associated with teaching styles that emphasise problem solving and homework. In addition, teachers seem to adapt their optimal teaching style to the student representing the 70th percentile on attainment.

In many cases pupil-level data is either not available or not accessible to the analyst. Thanassoulis (1996) puts forth an approach for ascertaining the presence, if any, and direction of differential school effectiveness using aggregate pupil data. The method in Thanassoulis (1996) is based on contrasting schools on the distribution of grades pupils obtain while allowing for the abilities of the pupils, for their family background and for the overall effectiveness in value-added of each school. The method is developed with reference to British secondary schools recruiting pupils at age 11 and measuring their exit achievements at age 16. It can, however, be adapted to other educational systems especially where the only difference is the grading system used and/or the ages of the pupils concerned.

The method in Thanassoulis (1996) begins with an assessment by DEA of the schools concerned on their effectiveness in value added. The set of input variables suggested to assess value added of secondary schools were the mean verbal reasoning score on entry and the percentage of students not receiving free school means (the latter being used as a surrogate for parental background of the pupil). The set of outputs used were the percentage of pupils placed after GCSEs (i.e. in work or further education), and the average exit GCSE score per pupil. GSCE is the General Certificate of Secondary Education which pupils obtained at that time in Britain at the completion of compulsory secondary education, normally at age 16. The number of A, B etc. grades achieved by the pupils of each school were publicly available (and defined here as NAj, NBj, etc. for school j). So it was possible to compute the aggregate GCSE score of school j as $\mathrm{Gj} = 8\mathrm{NAj} + 7\mathrm{NBj} + 6\mathrm{NCj}$

+ 5NDj + 4NEj + 3NFj + 2 NGj + NUj and thereby the mean GCSE score per pupil for each school. (The weights 8 for grade A, 7 for grade B etc. were at the time the accepted approach to converting letter grades to an aggregate numerical GCSE score).

A standard output oriented DEA model was used to compute the effectiveness of each school j_o, whose outputs were the efficiency score (θ^*) and the intensity variables (λ_j^*), the latter denoting how much a peer unit j contributes to the targets of the unit being assessed (j_o).

The grades profile of the efficient comparator "school", cj_o, of school j_o is given by:

$$PA_{cjo} = \sum_{j=1}^{n} \lambda_j^* PA_j$$
$$PB_{cjo} = \sum_{j=1}^{n} \lambda_j^* PB_j \tag{12.4}$$
$$PC_{cjo} = \sum_{j=1}^{n} \lambda_j^* PC_j$$

where PA_j is the number of A grades per pupil at school j, (computed as NA_j divided by the number of pupils in the school). PB_j … $\mathrm{PU_j}$ are defined in an analogous manner.

Thanassoulis (1996) suggests that "A comparison of the grades profiles of the efficient comparator cj_o and school j_o can be used to gauge the differential effectiveness of school j_o". The grades profiles of two schools can be compared in a number of ways. Thanassoulis (1996) suggests a simple way where grades A to C (top grades) are used to compute a component "ATOC" and the rest to compute a component "DTOU". The ATOC and DTOU component of school j are:

$$\mathrm{ATOC_j} = 8PA_j + 7PB_j + 6PC_j$$

and

$$\mathrm{DTOU_j} = 5PD_j + 4PE_j + 3PF_j + 2PG_j + PU_j.$$

Thanassoulis (1996) puts forth a procedure for computing the expected E(ATOC) and E(DTOU) components of school cj_o if it had had the same academic effectiveness as school j_o. (The details of how E(ATOC) and E(DTOU) can be computed are beyond the scope of this chapter but can be found in Thanassoulis (1996)). It is then suggested that:

– If ATOCj_o = E(ATOCcj_o) then we have no evidence of differential effectiveness between schools j_o and cj_o;

– If ATOC$j_o \neq$ E(ATOCcj_o) then:
  – if the difference is substantial, we have evidence that schools cj_o and j_o have dissimilar differential effectiveness over pupils of different academic ability;
  – the school with the larger ATOC component is likely to be more effective over the stronger (on entry) pupils.

"School" cj_o would reflect the grade profiles of the efficient peers to school j_o. Thus in effect the method looks for differential effectiveness between school j_o and its efficient peers. Clearly it is difficult to specify a general purpose threshold for the difference in ATOC components which would trigger an identification of dissimilar differential effectiveness between schools as this largely depends on the factors that might have been omitted from the input/output variables used in the DEA assessment. It is clear, however, that the larger the difference in the ATOC components of schools j_o and cj_o the stronger the indication of dissimilarity in differential effectiveness between school j_o and its efficient peers. Thus the method should identify at least the cases where there is substantial dissimilarity in differential effectiveness between schools.

If school j_o turns out to be efficient and self-comparator we cannot draw any conclusions about its differential effectiveness. Further, the comparative basis of the method outlined here means that it may fail to identify differential effectiveness in those cases where the schools being compared have similar differential effectiveness. On the other hand where the schools do differ they become good examples to one another on teaching practices, which can benefit those ability ranges their current teaching practices disadvantage. These difficulties are overcome by attempting to identify differential school effectiveness using pupil level data as outlined later in this chapter.

12.2.2.3 On-line Platforms for Assessment of Schools

Most DEA applications in education, whether using aggregate pupil or pupil-level data, normally rely on analysts, usually from outside the schools concerned, to conduct the assessments of VA. However, there are cases where the assessments are set up in such a manner that schools can self-assess on an on-going basis as new data becomes available. We outline here one such approach, operationalised in Portugal.

The Portuguese Education Ministry supplies every year the media with exam results and they publish school rankings based mostly on these results. Since 2013 the Ministry also provides some contextual variables that allow a contextualized ranking analysis. There is, however, a privately run and innovative platform called BESP (feg.porto.ucp.pt/besp) that provides information to the general public on a set of indicators based on the national exam databases (see Portela et al. 2011). This platform concerns only secondary schools (with students from the 10th year to the 12th year of schooling). It is designed to serve not only the general public, but also to serve schools as a tool for self-evaluation. Indeed, schools can enter data through on-line questionnaires and immediately have access to a number of graphs that show how they are evolving on certain indicators over time, as well as how they perform on specific indicators in relation to a comparable set of schools, which can be in the country, district, or of the same type, etc. A DEA procedure is embedded within BESP, which allows schools to choose a customized set of inputs and outputs and compare their performance with that of other schools.

The information displayed in BESP regarding individual indicators is illustrated in Fig. 12.3.

Graph (a) shows the distribution of the indicator 'average results on national exam', with the percentile for the selected school displayed. In the example in Fig. 12.3 the school chosen lies at the 73.39 percentile. In the radar graph (b) of

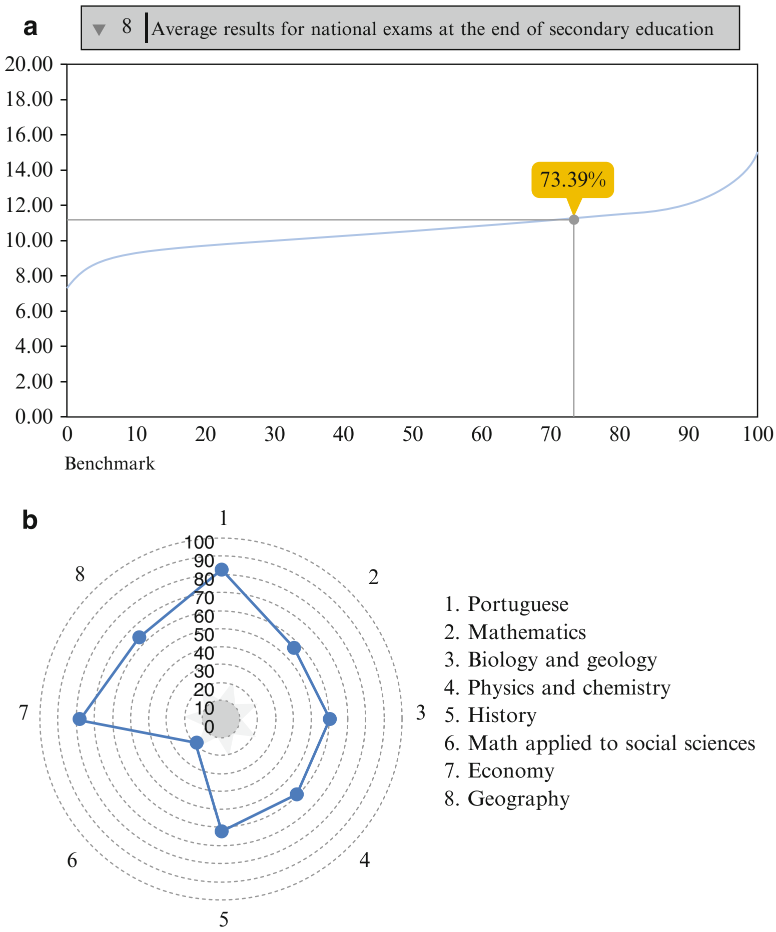


Fig. 12.3 BESP – results for the indicator average classification of the school in final exams (Picture taken from Portela et al. (2011))

Table 12.2 Set of inputs and outputs available within BESP

Inputs	Outputs
Average grades in Portuguese in years t-2 and t-3	Average grades in the kth national exams in year t
Average grades in Mathematics in years t-2 and t-3	Percentage of school students that took exam k
Parents' average years of schooling	Percentage of students concluding secondary education in the 'normal' 3 years
Economic context of the school (computed on the basis of the number of pupils in the school that receive subsidies from the state)	Percentage of students that proceeded to university
	Percentage of students that did not abandon the school

k = Portuguese, mathematics, biology and geology, physics and chemistry, history, economy, geography, mathematics for social sciences

Fig. 12.3 details are shown for the percentiles in which the school lies on each course that is included in the overall average for the school. The radar shows in which subjects the school has better percentile position (e.g. subjects 1 and 7) and the ones where the school has worse percentile position (subject 6).

BESP also enables schools to self-assess using DEA. A menu of potential inputs and outputs (see Table 12.2) are presented to the school from which it can choose a subset on which to be evaluated. The assessment is at the school level using aggregate pupil data.

BESP enables schools to solve on-line DEA models with the inputs and outputs selected from the above list. More details on how to conduct these assessments can be found in Portela et al. (2011, 2012). Given that not many schools upload data into the platform, the assessment possible at the time of writing is the one with the first two inputs and the first output in Table 12.2, for which data are available from public exam databases. Clearly this type of analysis focuses on academic outcomes and neglects other, important non-academic outcomes of education (such as developing inter-personal skills and nurturing responsible social skills).

Focusing on academic outcomes, if a school selects the available inputs and outputs in Table 12.2, the output displayed by BESP is the overall efficiency score and a radar showing how the observed inputs and outputs of the school compare to its potential attainments, controlling for its student intake as reflected in the input variables. In addition, radars showing how the school compares with its peers are also displayed, such that the school can identify the factors where other schools are doing better than itself. An example of such radars is shown in Fig. 12.4 for a school, which has an efficiency score of 79.2 %.

The peers (identified by the darker lines which in general enclose the lighter coloured lines) have similar or slightly higher inputs than the school assessed (the inputs are taken as aggregate grades on entry in this case – 'EntyGrades'). In spite of that, both peers achieve clearly higher average scores on exit in most of the courses. The above assessment shows the evaluation of the selected school against all schools in the country, but private schools should probably be excluded from the

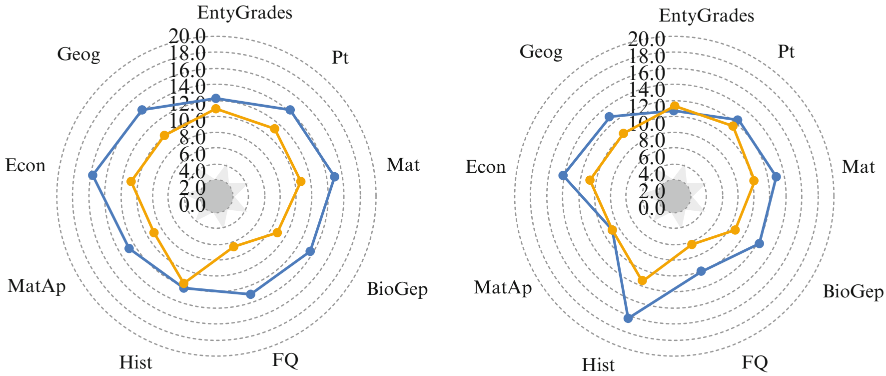


Fig. 12.4 A school being compared with its peers in BESP (Graph from Portela et al. 2011)

comparison set as the selected school is a public school. This can be readily done within BESP.

With the advent of the internet several internet-based tools such as the above have become available to publish information regarding school performance. The publication of results on school performance is normally the responsibility of national governments through their agencies. For example, in the UK, the Department for Children Schools and Families publishes performance tables (education.gov.uk/schools/performance/). The user can select any particular school and gain access to details on the school demographics as well as its performance on a number of indicators, including a measure of value added (VA). Also in the UK there is a web-based application called RAISE "report and analysis for improvement through school self-evaluation" (raiseonline.org), which enables schools to look at performance data in greater depth as part of their self-evaluation process. This application is, however, not available to the general public, but just to schools. This is also the case with the UK School Financial Benchmarking (SFB) website (sfb.teachernet.gov.uk) (for details see Ray 2006; Ray et al. 2009).

In Norway Skoleporten is a national school accountability system, which contains publicly available data on indicators for results, resource use and learning environment (details in Haegeland 2006). The Swedish National Agency for Education also publishes data for all levels of the education system (skolverket.se/sb/d/190). Apart from data on several different indicators, the agency also publishes expected results for each individual school, estimated using linear regression. In Portugal the Education Ministry has a website (infoescolas.mec.pt/) that shows educational statistics for all the schools, including some progress measures similar to VA measures.

In the US there is the Tennessee Value Added Assessment System (TVAAS) developed by Sanders et al. (1997), which has a public website (tn.gov/education/data/TVAAS.shtml) where the general public can access reports on VA. This system has also been adopted by other states, and in particular Pennsylvania, which has the PVAAS, (available at pvaas.sas.com) and Ohio, which has EVAAS (available at ohiova.sas.com).

Note that the trend towards internet-benchmarking platforms that allow on-line and immediate comparisons between production units can also be found in sectors beyond education. For example, the construction industry has a benchmarking platform icBench (icbench.net) (Costa et al. 2007) available in Portugal, whereas in the US there is BM&M (construction-institute.org), and in the UK there is KPIzone (kpizone.com). None of these or the above examples, however, use DEA to carry out performance assessments. An example of a platform that allows also for aggregate assessments through DEA can be found in iDEAs (isye.gatech.edu/ideas) (see Johnson and McGinnis 2011). This platform is targeted at warehouses or other industrial systems and combines benchmarking and DEA to allow managers to benchmark their performance against others. Bogetoft and Nielsen (2005) report an internet-based benchmarking system, applied to Danish commercial and savings banks, that incorporates a DEA model. The developed platform is currently being commercialized and applied to other industries (Ibensoft Aps 2013).

12.2.3 DEA Applications Using Pupil-Level Data

Given the hierarchical structure of education, multilevel models have been the main instrument of analysis in 'educational production function' approaches. To this popularity contributed, amongst others, the development and enhancement of hierarchical modelling techniques, known as **multilevel modelling**, by Goldstein (1987) and Raudenbush and Bryk (1986). Additional impetus was given by Sanders et al. (1997), who carried out a project in Tennessee (USA) which included not only the estimation of the VA of schools but also of teachers. More recent examples of applications using pupil level data can be found in Agasisti et al. (2014), Mancebón et al. (2012) and Hanushek et al. (2013).

There are not very many applications of DEA to pupil-level data. As we have seen, DEA models were initially mainly applied to aggregate (e.g. school level) data. To the authors' knowledge Thanassoulis (1999) is the first DEA study that used pupil-level data, to set achievement targets for pupils. School effectiveness or VA was not investigated in the Thanassoulis (1999) paper. The measurement of VA of schools through DEA, using pupil-level data, was first attempted by Portela and Thanassoulis (2001) and Thanassoulis and Portela (2002). The approach adopted by the authors to compute the VA of schools and the key findings will be detailed in the next two sections.[2]

[2] Note that parametric frontier models have also been widely used in the educational context, but these will not be detailed in this chapter (examples of pupil-level studies through stochastic frontier models can be seen amongst others in Cordero-Ferrera et al. (2011), Deutsch et al. (2013), Perelman and Santín (2011) or Crespo-Cebada et al. (2014)).

12.2.3.1 An Overview of DEA Applied to Pupil-Level Data

We will use the approach of Portela and Thanassoulis (2001) to outline the use of DEA with pupil-level data. Their approach was inspired by the approach outlined in Sect. 12.2.2.1 for disentangling managerial from program efficiency introduced originally by Charnes et al. (1981). The DEA approach mimics the parametric multilevel modelling approach in that pupils are assessed hierarchically so that pupil effects can be isolated from higher level effects such as school effects, local education authority effects etc. The approach is illustrated through Fig. 12.5, where crosses represent students from various schools and the circle-shaded crosses represent students from a particular target school. On the axes we depict attainment of the student on entry to a certain educational stage and attainment on exit from that same educational stage, respectively. These two variables normally feature in VA assessments in secondary education but they are only a subset of the input and output variables generally used. The data in Fig. 12.5 are real, corresponding to grades on entry (at the 9th grade) and grades on exit (at the 12th grade) of students from Portuguese secondary schools in a certain 3-year period. Pupils beyond the global frontier are deemed 'outlier', not permitted to define the frontier. (Two pupils of the target school (shaded crosses) are also deemed outlier.) The data were used within an evaluation programme called AVES (for details see Portela and Camanho 2010).

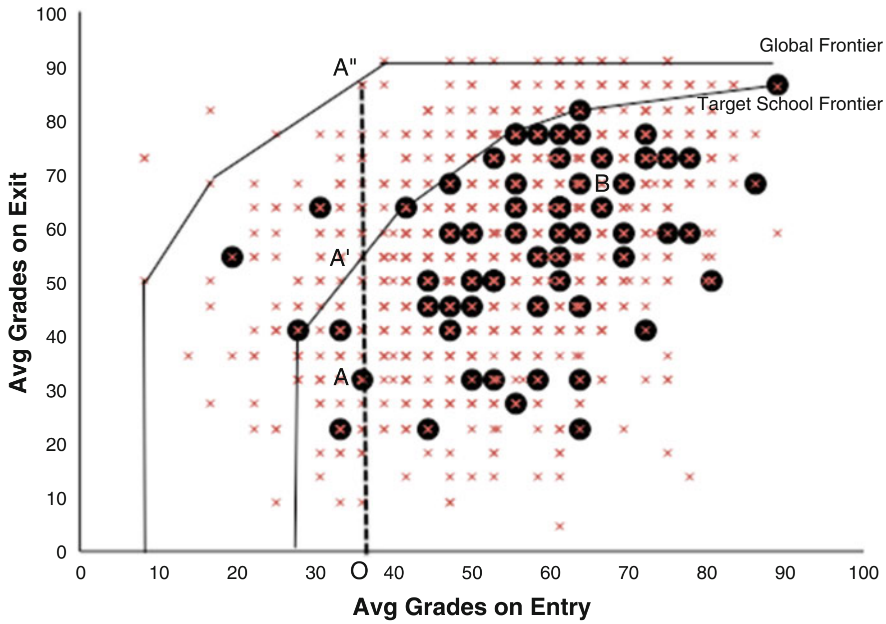


Fig. 12.5 Illustration of the DEA approach to assess VA using pupil-level data

Consider now pupil A attending the target school in Fig. 12.5. The efficiency of this student with reference to the frontier of students attending the same school can be measured through OA/OA′. If instead we compare the student to the overall set of students from all schools (global frontier in Fig. 12.5) the efficiency of student A is measured through the distance OA/OA″. As a result, the global efficiency score (or pupil-within-all-schools efficiency) decomposes into two components:

$$\frac{OA}{OA''} = \frac{OA}{OA'} \times \frac{OA'}{OA''} \tag{12.5}$$

A component that is attributed to the pupil, OA/OA′ (termed pupil-within-school efficiency in Portela and Thanassoulis (2001)), and a component attributable to the school OA′/OA‴ (termed school-within-all-schools efficiency in Portela and Thanassoulis (2001)). The pupil effect (OA/OA′) incorporates the shortfall in achievement of pupil A that is due to the pupil alone, as the target school he/she attends has demonstrated is able to place students with the entry attainments of pupil A at point A′ on exit. The school effect (OA′/OA″’) reflects the component of under attainment of pupil A that cannot be attributed to the pupil (imagining that he/she was at point A′) but to the school, that did not foster as much attainment on exit as other schools in the sample did, for students with entry attainment as that of pupil A.

The pupil-within-school efficiency scores reflect the diversity of achievements attributable to the pupils in that school. Within school efficiencies are not comparable across schools except to reflect relative diversity of attainment of pupils of different schools. For example, in a school where the mean of the pupil-within-school efficiency is lower than in another school the pupils of the former school will in general be further away from their school frontier than in the case of the latter school. However, we cannot say in which school we have higher value added in absolute terms as this will depend on the relative positioning of the school frontiers. In schools where pupils are generally close to maximum possible attainment we expect high average pupil-within-school efficiency scores, whereas a low average is expected for schools with highly heterogeneous pupil attainments.

The school-within-all-schools efficiency reflects the shortfall, if any, between the current best attainment of a student and the corresponding best exit attainment available across all schools for a given entry level.

In Fig. 12.5 we have considered just two levels of analysis, pupil and school. Originally in Portela and Thanassoulis (2001) three levels of analysis were considered the pupil, the school, and the school type (Comprehensive, Grant Maintained and Independent schools). An alternative 3-level analysis is one where pupils are level 1, classes are level 2 (assuming pupils are taught in different classes), and schools are level 3.

Teacher effects on pupil achievement is a recent issue addressed in the literature (see e.g. Chetty et al. 2014), and this effect can be estimated using pupil-level data in the manner outlined here, where a teacher is associated with a given class of pupils. To illustrate the approach to assessing teacher effectiveness let us refer to

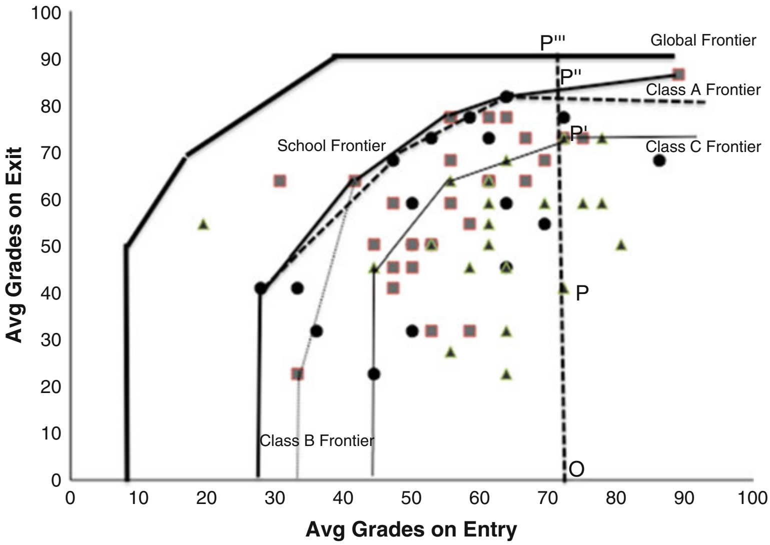


Fig. 12.6 Representing pupils by class within school

Fig. 12.6 where a global frontier has been drawn mapped out by attainments across all schools and their classes. Consider now a target school which has three classes (A, B, and C) for teaching the course under analysis. A frontier for each class can be constructed following the same principles as in Fig. 12.5 and distances computed accordingly. This is illustrated in Fig. 12.6 (where the all-schools (or global) frontier was kept, but students defining that frontier have been omitted).

Frontiers of class A (students from this class are represented by circular dots) and B (students represented by squares) are very close to the school frontier (constituted mainly by pupils from Class A and from Class B), whereas class C frontier (students represented by triangles) lags behind the overall school frontier. For pupil P attending Class C in the school under consideration we have the following decomposition of a global efficiency score (or pupil-within-all-schools efficiency):

$$\frac{OP}{OP'''} = \frac{OP}{OP'} \times \frac{OP'}{OP''} \times \frac{OP''}{OP'''} \tag{12.6}$$

The first term in the right hand side of (12.6) measures the pupil effect, the second term measures the classroom effect, and the third term measures the school effect, on a pupil having the entry attainment of pupil P. If attainment on a single subject is being considered (e.g. maths) the classroom effect is indeed the teacher effect assuming only one teacher teaches the subject concerned. In such a case the component (OP′/OP″) is a measure of pupil under-attainment that is attributable to the teacher of class C as other classes in the same school have shown that, for the initial level of attainment of pupil P, higher attainment than at P′ was possible.

If attainment on entry and on exit cover a varied number of subjects, the component OP′/OP″ will in fact be attributable to the class or more precisely to all teachers teaching the subjects under consideration to the class concerned.

The approach considered here can be generalised to account for other types of effects, depending on the number of categories under which pupils can be grouped beyond classes and schools – e.g. gender, ethnicity, country, etc. For example, country specific effects can be estimated if one assumes that inner frontiers in Fig. 12.6 represent schools, enveloped by a country frontier, and a global frontier enveloping all country frontiers. In fact the approach outlined can be seen as a more general one that can account for categorical variables in DEA assessments, which is originally designed to deal only with continuous variables.

Thieme et al. (2013) have extended the approach using pupil-level data outlined above by considering not only pupil-level variables but also school level variables. Accordingly they divided the school effect into three types: resource endowment effects, peer effects and selection bias effect. The first is related to the fact that pupils with different levels of performance can be placed in schools with different resource endowments; the second is related to positive externalities that can happen when students try to emulate their peers to reinforce their identification with the group; the third is related to the fact that more able and motivated students may place themselves into certain types of schools. The inclusion of school variables in the Thieme et al. (2013) approach works by defining additional frontiers in Fig. 12.5 or Fig. 12.6, where, for example to estimate the resource endowment effect, each pupil would be compared to its own school frontier, to a frontier including all schools that operate with no more resources than the pupil's school, and finally to the global (all-schools) frontier.

12.2.3.2 Applying DEA to Pupil-Level Data

In this section we briefly reflect some of the main practical aspects of using DEA on pupil level data.

Input–Output Variables

The selection of appropriate input and output variables is a fundamental step in any DEA analysis (Thanassoulis 2001, Chap. 5; Emrouznejad and De Witte 2010). The choice of inputs and outputs depends on the perspective from which the assessment is to be carried out. For example, in the case of aggregate level data, as we saw, adopting a perspective of school efficiency led to a different set of input output variables compared to a perspective of school effectiveness. Regarding pupil-level data one is more constrained regarding the inputs and outputs to consider as these may relate to pupils and not to schools (in spite of school variables being also possible to consider). As a result pupil-level analyses use in general two types of variables: categorical and continuous variables. The continuous variables typically include measures of attainment by pupils on entry as an input, and measures of attainment by pupils on exit as an output. Attainments are typically in numerical form e.g. mean grade across a set of subjects or grades by subject (e.g. maths, science, English etc.).

Other continuous variables may relate to age of pupil, or various measures of socio-economic and parental education background (see De Witte and López-Torres 2015 for relevant references). Categorical variables are those whose impact on pupil attainment is to be investigated such as class, school, type of school, gender, ethnicity, etc. As seen before, the most immediate way of considering these types of variables is through the consideration of various frontiers for different categories of the variables (e.g. girls vs males frontiers, school frontiers, classes frontiers, etc.). The literature on the use of categorical variables is also related with the literature on non-discretionary factors (which are in many instances categorical) and initiated with Banker and Morey (1986), and followed by Ruggiero (1998). A note of caution is required when data is subdivided following various categories, as one should assume that these divisions still allow a reasonable number of comparator units to be assessed.

Dealing with Outliers

One of the issues with using pupil-level data is that any chance events (e.g. fluke high or low attainment by a pupil on a subject) is not mitigated in the way it would be when using aggregate data across pupils. As in DEA results are highly influenced by observations which are especially efficient, it is customary to seek to identify 'outliers' on efficiency. In DEA outliers are those observations that have efficiency much higher than 100 % when they are assessed relative to a frontier drawn on all units, excluding the observation being tested for outlier status. Such observations can 'pull' the frontier to largely unattainable levels setting a biased benchmark for other observations to aspire to. Observations with very low efficiency on the other hand, which could be deemed outlier in a statistical sense, do not present a problem in DEA as they do not impact the referent frontier for any units other than the pupil itself (see a discussion in De Witte and Marques 2010).

There are a number of procedures available in the literature for identifying and dealing with outliers in DEA (e.g. see Thanassoulis et al. 2008, Sect. 3.6.4). Earlier methodologies (see De Witte and Marques 2010 for an overview) typically identified and eliminated outlying (super-efficient) data. By removing the outlying observations from the data, one risks losing also the most interesting observations. More recent approaches such as the robust order-m or order-alpha techniques mitigate the influence of outliers without removing them from the sample (see Sect. 12.4.2). It is clear that approaches of this kind for dealing with outliers are very attractive in assessing efficiency in education where outliers at pupil level can be often observed.

Obtaining Efficiency Estimates

The computation of efficiency scores to be included in decomposition (12.5) or (12.6) can be computed through traditional DEA models in the literature. The models typically assume variable returns to scale and are output oriented. The assumption of variable returns to scale is dictated by the fact that attainments on entry and exit are typically percentages or indices not expected to follow a strict mutual proportionality. Output orientation is sensible given that attainment on entry or socio-economic characteristics as inputs, are determined before entry to the stage of education under assessment.

The procedure for computing the efficiency estimates, used in Portela and Thanassoulis (2001) involves the following steps:

– *Assessing the efficiency of each pupil in relation to the global frontier to obtain the pupil-within-all-schools efficiency of the pupil;*
– *Assessing the efficiency of each pupil in relation to its own school frontier to obtain the pupil-within-school efficiency for the pupil.*

The school-within-all-schools efficiency (or VA) at the entry level of each pupil is obtained as the ratio of the pupil-within-all schools to the pupil-within-school efficiency of that pupil. It should be noted that though this approach in Portela and Thanassoulis (2001) is very close to that of program efficiency in Charnes et al. (1981), outlined earlier, it is not identical. The computational procedure of Charnes et al. (1981) if implemented would have implied the use of three rather than two steps. Step (1) would involve computing efficiency scores of pupils within schools as above; step (2) would imply replacing each pupil's outputs by its frontier targets outputs obtained from (1), including non-radial components; Step (3) would require assessing the efficiency of all pupil targets as derived in step (2) to obtain the school-within-all-schools efficiency which is termed 'program' efficiency in Charnes et al. (1981). For larger data sets the approach of Portela and Thanassoulis (2001) is computationally less demanding. The Portela and Thanassoulis (2001) approach captures distances between all frontier points of the school and the all-schools frontiers, whereas the Charnes et al. (1981) approach only uses the efficient part of the school frontier (but not so for the all schools frontier).

Other than classical DEA, which assumes a convex production possibility set, can also be used with pupil-level data. For example, De Witte et al. (2010) modelled the 'production function' as a Free Disposal Hull (FDH) technology. Their paper controlled for outliers using the order-m method of Cazals et al. (2002). See also Thieme et al. (2013) for a similar application.

Aggregation of Pupil Level Results

Although pupil-level scores are of interest *per se*, in pupil-level analyses we are also usually interested in assessing performance at the school or at the school district level. Views about performance at these more aggregate levels can be gained by aggregating in a suitable manner the pupil level efficiencies. Pupil-within-school efficiencies are normally aggregated in the form of a geometric mean. The same can be done for school-within-all-schools efficiencies. When geometric means are used in this way the means are decomposable as indicated earlier into pupil-within-school and school-within-all schools components. This does not hold true if aggregation is by means of arithmetic averages. In both cases, the average (either arithmetic or geometric) reflects the aggregate as long as there is no correlation between efficiency and output, which is pupil attainment on exit in our case (Karagiannis 2015). Otherwise, the aggregation rules proposed by Färe and Zelenyuk (2003) and Färe and Karagiannis (2014) should be applied.

Extracting Additional Information on Performance from Longitudinal Pupil-Level Data

The approach for using pupil-level data in assessing VA at schools as described so far is couched in terms of one period of time only. However, in practice the evolution of VA at a school over time is of utmost interest. In the context of a multi-period assessment of VA, we have an additional consideration in that the global frontier of each hierarchical level (school global frontier) would in general be different for each time period. For example the global frontier enveloping all school frontiers may change from one period to the next, and as a result even for a school whose frontier did not move from one period to the next it may still show say a decrease in VA if the global frontier has improved. One approach to overcoming this problem is proposed by Portela et al. (2013) who suggested the use a stable metafrontier defined by all data for all time periods rather than use an evolving global frontier over time. That is, for an assessment where data is available for the last 5 years, all 5 years of data would be considered in defining the metafrontier. The disadvantage of this approach is that the efficiencies relative to the stable metafrontier will reflect best performance observed over the entire long-term period rather than what might have been attainable only up to some earlier point in time under consideration. The significance of this drawback will depend on the use to which the findings of the analysis are to be put. One can always of course also compute comparative performance up to specific points in time if that is of interest as opposed to a global view of the long-term performance of pupils and schools.

Another drawback of this approach is the fact that as time passes, data from new periods become available, and the replication of the analysis would imply a different metafrontier (with all data observed until the period of analysis). However, in practical examples when a considerable number of time periods are included in the definition of the metafrontier, it is likely that it is approximately stable over time and new data can be added without provoking many changes to the frontier.[3]

[3] In education settings, where the variables used in the analysis are grades obtained in national exams, it is unlikely that many changes happen from year to year, except if the syllabus of the course changes. Therefore, when a reasonable number of time períods is included in constructing a meta-frontier it is unlikely that new time periods will imply big changes in that frontier.

To pursue with the computation of VA change, it is important to note that this can only be done at the school level, as pupils are not the same over different cycles of studies. That is, data usually relate to a cohort of pupils analysed at the end of a certain cycle of studies (accounting for their attainment on entry at that cycle). In the next period of time, the cohort for that same cycle of studies is different, as a new cohort is finishing the cycle. It is possible that the longitudinal analysis could track the same pupils over time and over different cycles of studies. In that case the evolution of the VA at the pupil-level would be possible to assess. However, to the authors' knowledge there is no DEA application focusing on such a longitudinal analysis. A good example of such type of analysis can be seen in the Tennessee value added system.

Portela et al. (2013) define Value-Added Change (VAC) at school level, as an aggregate of the VA scores computed at the entry levels of pupils corresponding to the exit cohorts concerned. Thus denoting VA_s^t the measure of VA of school s for the exiting cohort in period t, and va_{js}^t as a measure of VA of pupil j exiting school s in period t, we can aggregate through a geometric mean pupils' VA values to obtain a measure of the school VA. Thus value added change from period t to t + 1 denoted VAC(t,t + 1) can be defined as shown in (12.7).

$$VAC(t,t+1) = \frac{VA_s^{t+1}}{VA_s^t} = \frac{\left(\prod_j va_{js}^{t+1}\right)^{1/N_s^{t+1}}}{\left(\prod_j va_{js}^{t}\right)^{1/N_s^{t}}} \tag{12.7}$$

As mentioned before, VA is a measure of the distance between the school frontier and a referent frontier, and conveys no information regarding how students perform within the school at individual student level. Clearly both the school and the pupil-level measures are of interest for managing value added at a school for the better. Portela et al. (2013) compute a Malmquist index inspired measure to capture performance change over time at pupil-level. The measure is based on measures developed in Camanho and Dyson (2006) and Portela et al. (2011), and is of the following form:

$$M_s = \frac{\left(\prod_j E_p\left(X_j^{t+1}, Y_j^{t+1}\right)\right)^{1/N_s^{t+1}}}{\left(\prod_j E_P\left(X_j^{t}, Y_j^{t}\right)\right)^{1/N_s^{t}}} \tag{12.8}$$

where $E_P(X_j^t, Y_j^t)$ is the efficiency score of pupil j observed in period t as computed in relation to the frontier P (the pooled metafrontier).

This index can be decomposed into an efficiency change or catch up component and a frontier shift. As shown in Portela et al. (2013) the frontier shift component is precisely the value added change defined in (12.7), whereas the catch up component (CU) is given by:

$$CU_s = \frac{\left(\prod_j E_{t+1}^s\left(X_j^{t+1}, Y_j^{t+1}\right)\right)^{1/N_s^{t+1}}}{\left(\prod_j E_t^s\left(X_j^t, Y_j^t\right)\right)^{1/N_s^t}} \tag{12.9}$$

That is, the ratio of the average efficiency of pupils in period t + 1 in relation to the school frontier of that period, to the average efficiency of pupils in t in relation to that period's school frontier. It is recalled that CU_s can only convey the relative level of dispersion of efficiencies of pupils relative to the school frontier. That is, a value larger than 1 would imply that pupils in period t + 1 are closer to the frontier of that period than pupils in period t, on average. However no conclusion can be drawn as to whether they perform better or worse in period t + 1 compared to period t as this will depend on the relative positioning of the school frontiers in periods t and t + 1.

12.2.3.3 Putting to Use the Findings from DEA Assessments of Pupil-Level Data

Each DEA application has its own aims and the perspective used varies depending on the stakeholders in whose behalf the assessment is undertaken. We illustrate here some typical uses made of the results from DEA assessments.

Gaining Insights into Components of Performance Attributable to Pupils, Schools or Other Hierarchical Levels

An obvious initial output from the DEA analysis of pupil-level hierarchical data are the various efficiency scores. These are provided at the pupil-level, and therefore their analysis is possible at various levels of hierarchical or categorical aggregation that may apply. Graphs showing for each school the global efficiency scores (pupil-within-all-schools) and the pupil-within-school efficiency scores are of particular interest (Thanassoulis and Portela 2002; Portela and Camanho 2010; Thieme et al. 2013). Such graphs are shown in Fig. 12.7 (using a real data set) for two schools with different profiles.

School A and B show similar pupil-within-school efficiency scores (average of 79 % for school A and 77 % for school B). However, they show remarkable differences regarding the within-all-schools efficiency scores. In school B

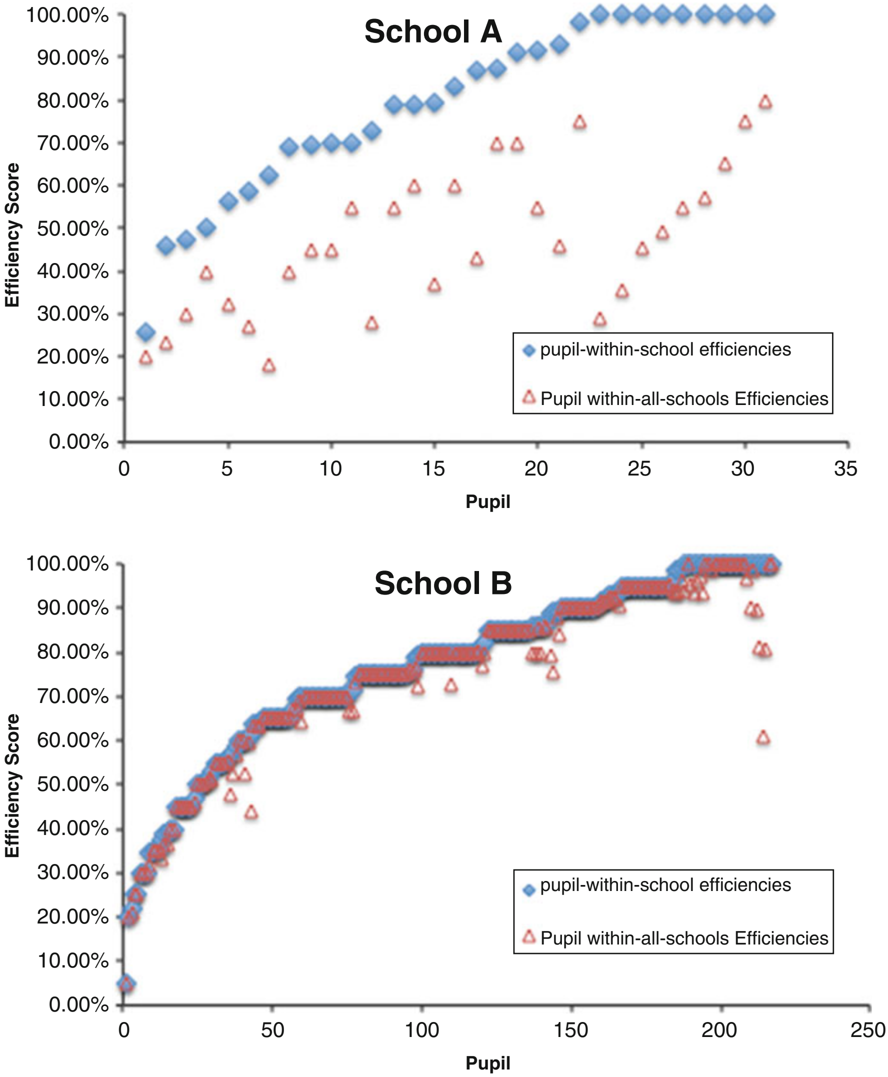


Fig. 12.7 Distribution of pupil-within-school and within-all-schools efficiency scores

pupil-within-all-schools efficiencies are similar to within school efficiencies (75 % on average) and in school A they are dramatically lower (47 % on average). The ratio of the pupil-within-all-schools to the pupil-within-school efficiency, it is recalled, captures the school effect or school VA on a pupil's performance and it is referred to as the school-within-all schools efficiency at the pupil data point. This ratio for school B has an average value of 98 % whereas for school A it has an average value of 60 %. This means that school A has a larger detrimental impact on pupil attainment than does school B.

Information of the foregoing type can be summarised in a tabular form contrasting the average school-within-all-schools efficiency with the average pupil-within-school efficiency for each school. In Portela and Camanho (2010) such tables have been compiled. Similar tables are also found in Portela et al. (2013), where change in school and pupil performance over time are also summarised.

Identifying Role Model Pupils and Setting Achievement Targets

An approach for target setting at pupil-level is detailed in Thanassoulis (1999). The approach is based on the use of DEA with data at two levels – pupil and school – but the approach can be extended to any number of levels. E.g. the classroom, the school, the global set of schools, etc. Targets are based on the projection of the attainments of the pupil to a frontier point, corresponding to the pupil's attainments on entry to the level of education concerned. These targets flow out directly when using DEA (e.g. the DEA software available from www.deasoftware.co.uk). The least challenging for the pupil targets would be those based on his/her within-school class frontier, if data by class have been analysed. Such targets should be attainable by the pupil if his/her teacher were to maintain his/her current level of effectiveness with pupils. Consider pupil P whose attainments to date are not on the efficient frontier. Role model pupils within the pupil P's class, with similar attainments on entry, are identified by the DEA model and their attainments to date can be used to establish the target attainments pupil P should be able to attain. Targets based on higher levels of aggregation of pupils follow a similar logic. For example, targets based on the global all-schools boundary can be set but would be more challenging for the pupil and they could involve a challenge to the school of the pupil too, if the school's within all schools efficiency is not 100 %. Role model pupils may also be harder to hold out as examples to an inefficient pupil if those pupils are from schools other than his/her own.

Except for Thanassoulis (1999) not many studies have explored target setting for pupils. However, personalised target setting is an important part of a school improvement process. Further, the approach also shows clearly the scope for further improvement in pupil attainment that rests on school rather than pupil efforts. When longitudinal data are available the static snapshot results can be complemented with projections on future performance, based on improvements in value added effectiveness in the past. This approach "enables schools and administrators to determine, given the expected growth rate of a particular group of students, what proportion of students will meet a desired standard, and this facilitates planning and resource allocation" (OECD 2008, p. 79). The Tennessee value added system provides projection reports where the trajectory of students is combined with the trajectory of the school's value added. These reports play an important role as "if a large number of students are projected to fall below the proficiency standard, the school has an early warning signal that it must aggressively address the factors... that are retarding student progress" (OECD 2008, p. 80).

Identification of Differential School Effectiveness

As mentioned earlier, early literature identified the issue of differential school effectiveness in the sense that the effectiveness of a school may differ by group of pupils (e.g. by ethnicity, social economic background, innate abilities, gender, etc.) (see e.g. OECD 2008). Differential effectiveness could be identified using aggregate data as we saw earlier, but pupil-level data lends itself much better for this purpose.

Take for example the case of Portuguese schools, assessed in Portela and Camanho (2010). If average VA is computed for each group of students according to 'innate –ability', one can have an idea of the profile of the school's effectiveness with different ability groups. Figure 12.8 illustrates this idea, with two schools, C and D, showing different profiles on their VA for different ability groups (represented on the horizontal-axis).

The VA of School C increases with rising pupil attainment on entry. The profile of this school is very close to that observed across the full set of schools under analysis. However, the school shows in all ability groups lower than average performance. In contrast, school D shows the reverse profile having its highest VA for pupils with the lowest attainment on entry, with VA falling as attainment on entry rises. This means that school D excels at making its weakest pupils on entry attain the best results on exit. This appears to be done at the cost of potentially not raising the attainment of the stronger on entry pupils as far as it might have been possible. Its VA is much below the average for the stronger on entry pupils. Note that in interpreting VA in the above manner care is needed to ensure the group concerned (e.g. of weak or strong pupils on entry) should be of sufficient size for the mean to be representative of the group (see for example Simpson 2005).

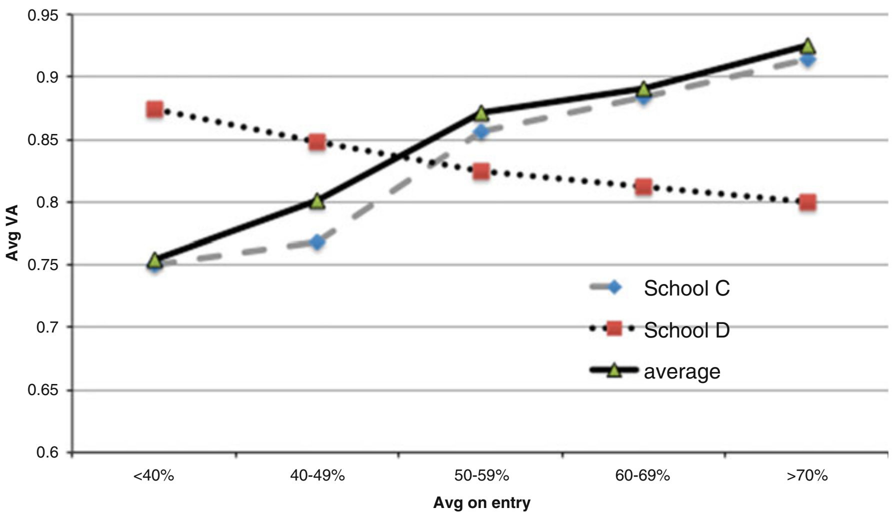


Fig. 12.8 VA of two schools by ability on entry group

The Influence of Environment on School Performance

Estimates of VA are of interest for parents to enable the choice of school for their children. On the other hand policy makers and school administrators have a different interest as it is important to "disentangle the contributions of the school context and the school practice to the gains of the students" (OECD 2008, p. 109). This interest at the policy level can be addressed using pupil-level data in a DEA framework. In particular, second stage procedures outlined earlier for explaining DEA results are a means to this end. There are various types of second stage analysis. One is through qualitative case studies where schools with very high VA and others with very low VA are scrutinised in detail through visits and interviews with school directors, teachers and other staff. For one approach of this type see e.g. Portela and Camanho (2007). Another approach used is a quantitative one where parametric techniques are deployed to regress the VA efficiency estimates on a number of possible explanatory variables. For example, Cherchye et al. (2010) have applied second stage models, but environmental variables were mainly defined at the pupil-level. The exception was the type of school. The type of school was also one of the school level variables used by Mancebón et al. (2012) but in this study several other school characteristics were included in a multilevel model (run in a first stage) to assess the science performance of Spanish students in PISA 2006. Most of the school level variables considered in the DEA assessment were contextual (e.g. proportion of students born in Spain, the proportion of girls in the school, the percentage of students not repeating any grade, average schooling years of mother, etc.). Only two variables concerned school practices: class size and ratio of computers per student. These types of variables are controllable and can be manipulated by policy makers to improve performance of schools. De Witte and Kortelainen (2013) analysed PISA 2006 scores of Dutch students, and performed a second stage analysis where they included a set of variables regarding pupils' background, but also school level variables (from which the contextual variables were the percentage of girls, school size, school autonomy, and average school socio-economic status). They used school practice variables in the form of minutes of maths lessons at the school and student-teacher ratio. Interestingly none of these school practice variables proved relevant in explaining effectiveness of schools. This is in line with previous literature (see the literature review by Hanushek 1986) which found little statistical evidence of the effect of school level variables on the effectiveness of schools.

Choosing Between DEA and Parametric Multilevel Modelling

As discussed in Sect. 12.2.3, multilevel models are the most applied methodology to pupil-level data. Multilevel models are regression-based models, ascertaining the variability in pupil attainment between different levels of aggregation of pupils. With two levels (the most frequent setting, e.g. level 1 is the pupil and level 2 the school) a dependent variable at pupil-level is regressed on a set of independent variables. The intercept is taken as a random variable for each school. Pictorially this is similar to imagining various regression lines, one for each school, where distances between observations and the regression line are taken as the pupil's random error (measuring pupils variability) and distances between school regression lines (or intercepts assuming they are parallel) are interpreted as the school's random error (measuring schools variability, or the school effect).

Some authors have compared DEA based models with multilevel model for evaluating schools effectiveness. For example, De Witte et al. (2010) used robust non-parametric FDH and parametric multilevel modelling on a sample of data relating of 3017 girls attending British single sex schools. They used a robust FDH frontier, and compared the results with those obtained from assessing the same girls using regression-based multilevel modelling. The aim was to contrast the pros and cons of the two approaches. Multilevel modelling is not geared to results at pupil-level. However, it does yield the proportions of variation in pupil attainment attributable to school versus the pupils and these could be compared with similar measures derived from DEA results. It is found the two approaches yield similar results in terms of the proportion of variation that can be attributed to the school and that which can be attributed to the pupil. Correlations found between pupil-schools and pupil-within-all-schools efficiencies are high and above 0.78.

The main disagreement between the two approaches lay on the estimate of school VA where the multilevel model finds a larger proportion of pupil under-attainment is attributed to the school compared to DEA. This is related to the fact that the DEA approach looks at how schools compare to each other when the comparison is based on their best attaining pupils (for their entry level) while multilevel modelling in essence captures the school component of pupil attainment by considering average levels of attainment for given entry level. Clearly both perspectives have their own advantages and disadvantages. Methods based on averages have the problem of sensitivity to exceptionally well or poorly performing pupils (for their entry level). That is, a cohort with some students performing well above or for that matter well below expectation may have undue influence on mean values. On the other hand, the comparison between frontiers as in DEA can be less prone to exceptionally good or poor performance provided care is taken to exclude prior to the DEA assessment outlier observations as discussed earlier. This, however, could be underestimating school impacts across the full body of pupils if the school has differential effectiveness (see Sect. 12.2.2.2) and only facilitates best performance for a small subset of pupils.

Given the fact that the information provided by the two methods is based on alternative (average versus 'efficient') views of performance, DEA and multilevel modelling should be seen as complementary rather than alternative methods (De Witte et al. 2010). This is also the conclusion arrived at in Portela and Camanho (2010), when looking at the differences between the application of the above DEA approach and the approach that at the time was being applied by the UK Department for Children School and Families in constructing public league tables on the VA of schools. Other studies have applied both methodologies, but not necessarily to compare them. For example Mancebón et al. (2012) used the two methods in a complementary manner, where multilevel models were used to identify the underlying production function, and then a DEA model was constructed, based on the evidence of the relevant variables in the multilevel model, to compare schools. The interest was mainly in a comparison between publicly funded and fee charging schools.

12.3 Applications of DEA in Higher Education: Institution Level

Education at any level involves the augmentation of human capital. But there are some key differences between the various levels of education. Primary education is typically delivered by generalist teachers, while secondary education involves specialist teachers. Nevertheless, at least at lower secondary level, the curriculum is broad and this imposes a degree of similarity of experience of students at different schools. At higher education level, the experience is considerably more specialised, and the distribution of subject specialisms varies more across institutions.

There are some additional features of higher education that serve to distinguish it from primary and secondary schooling, and these might make the sector intensely competitive and hence have implications for efficiency. For example, in most countries education is compulsory for students up to some level of secondary, but participation in higher education is optional. In addition, primary and secondary education is typically financed through the tax system, but students in higher education in many countries have to pay (often substantial) tuition fees. Finally, higher education is often undertaken during the early years of adulthood, and so students are no longer geographically constrained: many students choose to study at a location that is distant from their parental home.

Unlike staff in secondary or primary education, most if not all academics employed in higher education institutions (HEIs), are contractually expected to undertake research. Hence specialist teaching represents only some of the output of HEIs. Within each subject area, institutions also produce research and engage in knowledge transfer activities – that is, they first create new knowledge, and then they work with various organisations to exploit it.

All of the above considerations throw into sharp focus the need to consider, in any evaluation of higher education efficiency, the nature of providers in this sector as multi-output organisations. This, alongside the availability of good quality data, has made the higher education sector an important test bed for empirical applications of frontier methods generally and DEA in particular. DEA has been employed to assess cost efficiency and technical efficiency. Network DEA can offer insights into the production process and provide more detail to managers and policy-makers on how to improve efficiency.

Table 12.3 Set of inputs and outputs used in Thanassoulis et al. (2011)

Input	Outputs
Total operating cost (measured at constant prices and including depreciation, but excluding catering and student accommodation)	FTE undergraduate numbers in medicine
	FTE undergraduate numbers in sciences
	FTE undergraduate numbers in non-sciences
	FTE postgraduate numbers
	Research funding (quality-related funding council grants and other research grants, measured at constant prices)
	'Third mission' or 'knowledge transfer' activity, measured by income from other services rendered (again, at constant prices)

Note: FTE represents full-time equivalent

12.3.1 Assessments of Cost Efficiency in Higher Education Institutions Using DEA

Assessments of cost efficiency provide potentially useful information on economies of scale and scope and therefore can inform decisions on, for example, how the higher education sector might best be expanded, as well as on how HEIs might become more cost efficient.

An analysis of costs and efficiency in English higher education institutions is provided by Thanassoulis et al. (2011).[4] This analysis includes an input-oriented variable returns to scale DEA model using 3 years of data (2000–01 through 2002–3). The input-output variables are shown in Table 12.3.

Data are sourced from the Higher Education Statistics Agency and cover some 121 English HEIs. These institutions vary considerably in nature – from the ancient universities of Oxford and Cambridge through to civic universities and the new universities of the 1960s to institutions that received university status after 1992. In some institutions costs are distorted by the presence of substantial medical facilities, while in others they are not. The analysis is therefore undertaken both by considering the sample as a whole and separately within four distinct groups:

– pre-1992 institutions with medicine;
– pre-1992 institutions without medicine;
– institutions that were granted university status around 1992 (typically former polytechnics); and
– institutions that have gained university status more recently (typically former colleges of higher education, often affiliated with the GuildHE mission group).

[4] This follows earlier work by, for example, Athanassopoulos and Shale (1997) and Johnes (1998, 1999).

In conducting the DEA, a small number of outliers was identified and excluded from subsequent analysis.[5] The mean DEA technical efficiency of the sample (pooled over time and type of institutions but excluding outliers) was estimated to be about 86 %. This figure appears to be quite high, reflecting the competitive pressures that exist in this sector. The observation at the first quartile had an efficiency score of 79.3 %, again suggesting that inefficiencies are largely competed away. There was, nonetheless, a tail – the minimum efficiency observed was 27.5 %. This likely reflects heterogeneity in the sample of institutions. In particular, small, specialist institutions are likely to have costs that are high in relation to their outputs. This underlines the importance of conducting a more disaggregated analysis.

DEA is applied separately to institutions in each of the four subgroups identified above. Given that this essentially involves estimating a separate frontier for each group, comparisons across groups are valid only in terms of measures of relative homogeneity of efficiency. Mean efficiencies varied considerably across groups and was lowest for institutions that have gained university status more recently. Within this last group there was a tail of institutions where efficiency was below 0.3. This likely reflects heterogeneity, since this group comprises a particularly wide diversity of institutions – from specialist agricultural and arts colleges to generalist universities. Results for this category of institutions should therefore be treated with an appropriate degree of caution.

The DEA model assumed variable returns to scale. Institutions that have increasing, constant, or decreasing returns to scale were identified using the DEA models applied to separate sub-groups of institutions. Most HEIs have constant or decreasing returns to scale. This gave HEIs further information as to how they might be able to change scale size in order to exploit economies of scale. We return to this point below.

In order to estimate an 'efficient' unit cost for each type of output one of the approaches proposed in Thanassoulis (1996) termed 'DEA-RA' (DEA followed by Regression Analysis) was deployed. The purpose of this approach is to obtain estimated parameters for the efficiency frontier. This is done as follows: DEA is performed to identify the efficient and inefficient HEIs. Inefficient HEIs are then projected on to the Pareto efficient frontier by estimating efficient output levels for them using an output oriented radial DEA model, adding any slacks to radial projections. Finally, total operating cost (in pounds sterling) was regressed against the vector of 'efficient' output levels to derive an estimated linear cost function. For the full sample of institutions, the cost equation estimated was given by

$$\begin{array}{lllllll} C = & 13121X_m & + 5657X_s & + 4638X_a & + 3829X_p & + 1376R & + 1537K \\ & (12.0) & (19.6) & (18.9) & (7.1) & (84.9) & (14.4) \end{array} \tag{12.10}$$

[5] Following Thanassoulis (1999), the outliers have super-efficiencies (Andersen and Petersen 1993) in excess of 100 % and there is at least a 10 percentage point gap between the super-efficiencies of the outliers and the other observations.

where t-statistics appear in parentheses.[6] Here C denotes costs, X_m, X_s, and X_a denote respectively the output of undergraduates in medicine, sciences, and non-science disciplines, X_p is the output of postgraduates, and R and K denote respectively research income and knowledge transfer as defined in the outputs above. This specification of the cost function assumes no fixed costs. Thus in effect the estimated expression is a linear approximation to the CRS part of the piecewise linear VRS frontier.

The unit output costs derived are reasonable and in considerable accord with the unit costs of the same outputs estimated by a quadratic cost function using the same data in Johnes et al. (2005); they indicate that, at undergraduate level, medical education is the most costly to provide at just over £13,000 (or about US$20,000) per student. This is followed by tuition in the other sciences. Postgraduate education at £3829 per student is estimated to be, on average, less costly to provide than undergraduate education. DEA is not likely to have a very accurate picture here as, more than at UG level, institutions offer very diverse postgraduate courses ranging from expensive MBA degrees to much cheaper PhD degrees. The latter are likely to lower the estimated cost per postgraduate student further because many research postgraduates engage in both undergraduate teaching and joint research activity with staff; while the one-to-one supervision that such postgraduates receive is resource-intensive, their activities also serve to reduce the costs associated with undergraduate provision and with research.

Similar equations showing the relationship at the efficiency frontier between costs and outputs are derived by Thanassoulis et al. (2011) for each of the groups of institutions defined above. The broad picture, with medical education being the most costly, followed by other sciences, is replicated across all groups. The costs associated with postgraduate education vary markedly across different types of institution, however, this being most costly in pre-1992 institutions without medical schools.

Broadly comparable results are reported using a suite of alternative, parametric, estimation strategies, using as data the full sample of institutions. These include stochastic frontier, random effects, and generalised estimating equations to evaluate the parameters of a quadratic cost function, from which average incremental costs associated with each output type are calculated (Baumol et al. 1982). Whatever method is used, the cost associated with undergraduate tuition is highest for medicine, followed by the other sciences. In each of these statistical methods, however, the average incremental costs associated with postgraduate provision is estimated as being higher than that associated with undergraduate provision (other than in medicine). DEA is likely to give a better estimate of postgraduate cost per student once the sample is subdivided into four types of HEI as postgraduate provision is now more homogeneous within each sub-group.

[6] The t statistics should be treated with caution. They are high because the regression fits a line through a scatterplot that comprises observations that lie perfectly on piecewise linear segments.

By pooling the data over the 3 years, Thanassoulis et al. (2011) also investigate change over time in both the frontier and the position of each institution relative to that frontier. This is done using the Malmquist index approach.[7] It is established that, over the period from 2000–1 through 2002–3, the Malmquist index (measuring total factor productivity) shifted very little for pre-1992 universities without medical schools and for post-1992 universities. This index declined quite markedly in the other two groups, however, the median institution suffering a 6 % drop in productivity. Decomposing this change into the components due to shift of the frontier and changes in efficiency of individual units indicates that the decline is *all* due to a shifting frontier. The authors note that this may be an artefact of the data. To be specific, over this period prices associated with the purchases of higher education institutions tended to be rising more quickly than is indicated by general price inflation; consequently the data used for real operating costs may overestimate the real value of inputs in the later years of the study. It is not clear, however, why this would affect some types of university but not others.

Another aspect investigated by Thanassoulis et al. (2011) was the possible augmentation of output levels, notably student numbers, that would be feasible at current levels of expenditure if inefficiencies were to be eliminated. They did this using the output oriented DEA model in two ways. Firstly the model was used in its classical format which scales all outputs equiproportionately maintaining the mix of all outputs (students, research and third mission) in order to gain Pareto efficiency. The potential output augmentations based on this model showed that across the sector there was scope for about 10 % rise in undergraduate science, 15 % in non-science undergraduates and 17 % in postgraduate student numbers. About two thirds of these gains were possible through the elimination of technical inefficiency and the remainder through the additional elimination of scale inefficiencies (i.e. exploiting economies of scale). Looking at the different types of institution the largest rise in student numbers possible in relative terms was at higher education colleges ranging from 20 % for undergraduate science to 36 % for postgraduate students through a combination of scale and technical efficiency gains.

A second variant of the DEA model that Thanassoulis et al. (2011) used involved varying the priorities for output expansion so that only student number augmentations are used to gain efficiency. There were significant differences between these and the preceding results when priorities were uniform across all outputs. They report that when both technical and scale inefficiencies had been eliminated the percentage rise in science undergraduates doubled from 11 % to 22 % and there was a 10 percentage point rise in the number of postgraduate students from 17.52 % to 27.16 %. The least change was in undergraduate non-science students where the percentage gain rose from 15.26 % to 19.81 %. These were large potential gains because the model is such that it seeks for each HEI to raise those student numbers

[7] The index developed by Malmquist (1953) was adapted for use in a DEA context by several researchers in the 1990s. See, for example, Førsund (1993) and Färe and Grosskopf (1996a).

where the maximum gain in absolute terms can be made, unconstrained by the need to maintain the mix of outputs. In some cases the model suggested only one type of student be augmented (e.g. at one university the numbers only of science students rise), because that is where the maximum potential for gain in student numbers lies within given resource levels. In this sense the results represent the potential for gains not only by eliminating scale and technical inefficiency, but also eliminating '**allocative**' inefficiency in the sense of maximising aggregate student numbers by altering the mix of students where appropriate. The authors do, however, sound a note of caution as the model may be overestimating potential gains as the four categories of students used are not sufficiently uniform within each category and so DEA by its nature would base results on those institutions which have the 'cheapest' type of student within each category (e.g. there may be a substantial cost differential between educating say mathematics and biology students yet the model treats both types as simply science students).

The data set used by Thanassoulis et al. (2011) has been used to derive further results by Johnes et al. (2008). This work focuses on statistical approaches and includes consideration of stochastic frontier methods that allow evaluation of efficiency scores while estimating parametric cost functions. This work is usefully considered alongside non-parametric approaches such as DEA. The statistical approach, pioneered by Aigner et al. (1977), has, like DEA, its origins in the work of Farrell (1957), but rather than using linear programming to find the frontier it employs a variant of regression analysis in which the unexplained residual term is defined to include a non-normally distributed component due to inefficiency. By taking this approach, the full toolkit of statistical inference becomes available. The results obtained by Johnes et al. (2008) are broadly in line with those produced by DEA and discussed earlier – efficiency scores obtained using the different methods are positively correlated (though the correlation is not particularly strong). The study is notable for its attempts to include location and the quality of student intake as determinants of costs, though neither appears to be statistically significant.

Johnes and Johnes (2013) provide an update of these statistical frontier analyses, and, using panel data, allow for heterogeneity across institutions by using the latent class variant of the stochastic frontier model (Lazarsfeld and Henry 1968; Orea and Kumbhakar 2004; Greene 2005). The results are broadly supportive of earlier studies.[8] A more refined method that can be used to accommodate heterogeneity is the random parameter stochastic frontier model (Tsionas 2002; Greene 2005), and this is used in another study by Johnes and Johnes (2009). Once again, the qualitative nature of the results confirms the findings of other studies. Broadly

[8] We should note, however, that, when the panel is broken into several sub-periods and models estimated on each sub-period separately, the magnitude of some parameters varies widely across sub-periods suggesting that the results should be treated with caution. Moreover, the latent classes determined by the data are puzzling: one might expect a priori that each class would comprise HEIs with common characteristics (perhaps with research intensive institutions, and other institutions in another). But this is not the case, and the common factor relating the HEIs in a group is not obvious.

speaking, as more allowance is made for inter-institutional heterogeneity, the efficiency score attached to the typical institution increases, though outliers at the bottom end remain.

This raises an important conceptual issue surrounding the evaluation of efficiency. Some institutions produce a given vector of outputs at a higher cost than other institutions for quite legitimate reasons. For example, the ancient universities have real estate that is expensive to maintain and that may be less than ideally suited for purpose; their costs are therefore high relative to those of other institutions. This should not be considered a reflection of inefficiency, as these universities are providing a wider service to society through the maintenance of architectural heritage. Now there may be any number of factors of this kind that explain higher costs in one institution than another. Whether any one of these factors is legitimate or not – and hence whether the higher costs are due to inefficiency or not – is essentially a judgement call. While DEA and other frontier methods produce output that may be interpreted as measures of efficiency, there is always scope for debate about what exactly this output means.

12.3.2 Assessment of Technical Efficiency in Higher Education Institutions Using DEA

The cost function approach of the previous section assumes that firms wish to minimise costs (a potentially dubious assumption in the context of a not-for-profit sector such as higher education). Technical efficiency provides an indication of how well (efficiently) HEIs are using their physical inputs to produce outputs. DEA can be used to estimate output distance functions and hence technical efficiency in this context. While most of the studies which examine technical efficiency are at the level of the HEI, DEA can also be applied equally to data at student level. This compares with the assessment of secondary schools using pupil-level data as described in Sect. 12.2.3. Such student-level studies can be useful in disentangling the effects of HEI efficiency from that of a student's effectiveness (Johnes 2006b, c). This type of information is useful for choosing a strategy for improving both institutional and student value added.

Johnes (2008) provides an example of an output distance function for higher education estimated using DEA.[9] Staff (both academic and administrative), students (both undergraduate and postgraduate) and expenditure on academic services are the inputs into the process which produces teaching (graduates from

[9] Earlier studies using DEA to estimate output distance functions for higher education include Athanassopoulos and Shale (1997), Flegg et al. (2004) and Johnes (2006a). The last is noteworthy for its pioneering application of statistical tests for comparing nested DEA models (Pastor et al. 2002) and for testing for differences in production frontiers of distinct groups of DMUs (Charnes et al. 1981).

undergraduate and postgraduate programmes) and research (income for research purposes). A Malmquist index productivity analysis finds that productivity has grown (on average) by 1 % per annum over the period 1996/96 to 2004/05 and that this is a consequence of improvements in technology that have outweighed decreases in technical efficiency. Rapid changes in the higher education sector over the study period (such as growth in student numbers and the use of online support materials for example, routine use of online multiple choice questions and virtual learning environments) appear to have had a positive effect on the technology of production (pushing the frontier outwards) but this has been achieved at the expense of lower technical efficiency (as inefficient HEIs have struggled to keep up with best-practice performance).

One problem with these results is that they are based on a set of inputs and outputs which do not incorporate quality of student intake and of exit qualifications. A more recent study which attempts to address this problem (at least in terms of undergraduate teaching inputs and outputs) focuses on the effects on efficiency of mergers (Johnes 2014). Undergraduate student numbers are adjusted by entry qualification while graduates from undergraduate programmes are adjusted by category of degree result. The remaining inputs and outputs are as in the earlier study. An output-oriented DEA is applied to an unbalanced panel data set from 1996/97 to 2008/09. The sample is unbalanced for a number of reasons. First, some HEIs merged during the study period. Following merger the new institution was treated as a different entity from the HEIs which merged to form it. In addition, some HEIs entered the data base[10] during the period.

The results of applying DEA to the pooled data set indicate that technical efficiency across the sector is around 80 % (similar to estimates of cost efficiency). The study also makes a preliminary examination of the effect on efficiency of merger activity. HEIs are identified as pre-merging (those institutions which will merge at some stage in the study period), post-merger (those institutions formed from unions of others) and non-merging. The DEA results suggest that post-merger HEIs are typically more efficient than either pre- or non-merging HEIs. These broad conclusions are confirmed using parametric techniques. It is worthy of note, however, that the underlying characteristics of pre-, post- and non-merging HEIs are very different and so the observed efficiency differences could be a consequence of something other than merger. Moreover, a closer examination of the individual mergers indicates that while *mean* efficiency is higher following merger, the efficiency effects can vary by case and there are both winners and losers in the merging process.

Some recent work on efficiency in higher education has focused upon international comparisons. Agasisti and Johnes (2009), for example, use (both constant and output-oriented variable returns to scale) DEA models to compare the performance of institutions in Italy and England over the period between 2002–3 and 2004–5. This analysis employs a rich set of input variables, with data on the student

[10] Data were obtained from the Higher Education Statistics Agency (HESA).

intake, staff, and financial resources; as outputs, numbers of graduates at various levels and a measure of research activity are used. The analysis is conducted both by running separate DEA exercises for the two countries and – as a distinct exercise – running a DEA on the data combined across countries. From the latter analysis, it is established that technical efficiency measures are typically lower in Italian institutions than in their English counterparts; the mean technical efficiency for Italian institutions is just 64 %, compared with a mean score of 81 % in England. In the country-specific analyses, the mean efficiency of institutions is virtually identical in England and Italy, suggesting that the efficiency differences observed across the two countries are primarily attributable to country level effects. Meanwhile analysis of the Malmquist indices, suggests that the Italian institutions are closing the gap. While little change in total factor productivity is observed in English institutions over this period, average efficiency of Italian institutions increased.[11] This finding is in line with the characteristic catching up process whereby less efficient institutions learn good practice from their peers.

12.3.3 Assessment of Research Performance of Higher Education Institutions Using DEA

12.3.3.1 Identifying and Measuring Inputs and Outputs

Research activity may be viewed as a multi-input, multi-output production process with execution time that notably differs across disciplines and even in fields within disciplines. It involves several forms of human (e.g., academic staff, PhD students, research assistants), tangible (e.g., scientific instruments, materials) and intangible (e.g., accumulated knowledge, social networks) resources that are combined to produce an output called "new knowledge", which has also tangible (e.g., publications, patents, conference presentations) and intangible (e.g., tacit knowledge, consulting services) features. Besides these, research output has two other aspects that are of special interest in assessment exercises: quality (i.e., research excellence) and a value or impact, with the latter being measured by citations counts when academic impact is concerned and judgmentally when impact is in non academic (e.g. business or government) domains (e.g. in the UK Research Excellence Framework of 2013 non academic impact of research was given a weight of 20 % (http://www.thinkwrite.biz/pdfs/quick_impact.pdf) compared to 65 % for conventional research output such as journal articles). The decision-making units behind research activity vary depending on the level of analysis from individual researchers to institutions, such as departments, schools, or even the university as a whole.

[11] The productivity of institutions on the frontier in Italy slipped back over this time period, but the gain in efficiency of other institutions more than compensated for this, yielding an average efficiency increase across the country of a little under 10 %.

The foregoing make research assessment exercises quite a complicated task requiring several assumptions and simplifications to be made at the outset. The first of them concerns the length of the assessment period considered. This is clearly related to the length of the publication period. Both the period from a paper's date of submission to a journal and its acceptance and the period from acceptance to actual publication date differ even within the same discipline. This is due to among other factors the procedures followed by different journals (e.g., number of referees, review rounds, etc.). The shorter the assessment period considered the higher the likelihood that research performance measures will be affected by random factors. This is particularly true for evaluation exercises conducted at the individual researcher level and less for more aggregate levels of analysis, i.e., departments, schools, or universities. Even though there are no a priori norms, empirical evidence from bibliometric studies (Abramo et al. 2012c) suggest that the preferable assessment period is between 3 and 5 years, depending upon the academic discipline considered.

On the input side, measurement of production factors other than labour is in most of the cases difficult or even impossible due to lack of data. We thus usually assume that resources (i.e., scientific instruments, materials, etc.) available to the evaluated units are the same at least within the same field and within a given institution. In addition, we assume, unless data are available, that the hours available for research are the same for each individual in a given field category. This is a reasonable assumption for higher education systems where hours devoted to teaching are established by national regulations and are the same for all, regardless of academic rank. In this case, research can be evaluated separately from teaching as labour input is allocated between research and teaching in fixed proportions. It is a less reasonable assumption for higher education systems where there is a trade-off between research, teaching, and administrative tasks and this should explicitly be taken into account in the assessment exercise.

However, the cost of time is different and is reflected in the labour cost that varies across academic ranks. Since salaries of full, associate and assistant professors differ it would be appropriate to distinguish between them by including three different "types" of labour in the assessment exercise or to measure labour input by its cost if information on individual salaries is available. The purpose of this is to distinguish between different degrees of quality among the employed human resources. Using uniform labour input instead of labour cost will normally have a more severe impact at the individual researcher rather than at more aggregate (i.e., department, school or university) level.[12] For evaluation studies at the individual researcher level when information on salaries is not available or salaries

[12] Since available data show that more senior academic staff have more, better and highly valued (cited) publications, department or university rankings based on uniform labour input will favour units with greater concentration at higher academic ranks.

are equal within academic ranks as in some higher education systems (e.g., Italy, Greece), the second best option is to evaluate research productivity by academic rank (Abramo et al. 2013a). Nevertheless, empirical evidence from a recent bibliometric study by Abramo et al. (2010b) indicate that the effect of switching from uniform labour input to cost of labour seems to be minimal expect for outliers.

On the output side, as the intangible counterpart of research output is hard to measure, we consider only codified new knowledge in assessment exercises. These include articles in academic journals, research monographs, patents awards, and presentations in conferences, and their relative importance that differs by subject category and/or discipline. The most prevalent form of codification for research output is publications in academic journals, which is considered as an acceptable approximation of research output in many fields but less so in the arts, humanities and a good part of social sciences (Abramo et al. 2014). But as patents are often followed by publications that describe their content in the academic area and conference presentations usually precede publication of academic work, consideration of the number of publications alone to approximate research output may actually avoid in many cases a potential double counting. Publications may be further distinguished by the type of outlet where they are published into academic journals, chapters in edited books or proceedings, research monographs, reports, theses, etc.

The way we count publications may induce biases in performance evaluation as the number of co-authors as well as the quality/prestige of the publication outlet are factors that one may want to control for in measuring research output.[13] Following Lin et al. (2013) and Hagen (2014)[14] there are four counting methods for collaborative papers: whole counting, where each collaborating author receives full credit; straight counting, where the most prominent collaborator (being either the first author or the corresponding author) receives full credit and the rest receive none; fractional counting, where credit is shared either equally by the collaborators (simple fractional measure) or based on some predetermined weights (full fractional measure); (Simple fractional counting is appropriate when authors are listed in alphabetical order or when it is explicitly stated that authorship is equally shared. Full fractional counting may be based on weights provided by field experts or some other authority.) harmonic counting, where credit is determined by the formula $\frac{\left(\frac{1}{i}\right)}{\left(1+\left(\frac{1}{2}\right)+\left(\frac{1}{3}\right)+\ldots+\left(\frac{1}{N}\right)\right)}$ with i being the position of an author in the by line and

[13] A priori the quality of a publication is independent of the number of collaborators and thus we have to adjust publications counts by both factors.

[14] Hagen (2014) also provided the corresponding formula for harmonic counting in fields like medicine where senior authorship is usually assigned to the first and last collaborator, who are respectively the leader of the specific research and the leader of the entire research group.

N the number of collaborators.[15] For arguments in favour of or against each counting method see Hagen (2014), Lin et al. (2013) or (Abramo et al. 2013b).[16]

In addition, publications counting and accreditation also depend on the level of aggregation at which the assessment is conducted. The research output at the department level is equal to the sum of publications with at least one author belonging to this particular department. Note however, that a publication co-authored by researchers of the same department or university is considered only once in research assessment at the department/university level but it accounts for a particular fraction in individual level evaluations. Similarly, a publication co-authored by researchers from different departments of the same university, will be considered only once in the evaluation of the university but it will account for a particular fraction in assessment exercises at the department or individual level. However, a publication co-authored by researchers from different universities will account for a particular fraction even for university level evaluations in addition to its fractional contribution for department or individual level evaluations.

Turning to research output quality two alternative measures have been proposed in the literature: the journal's impact factor and articles citation counts. Even though many will argue in favour of the latter, the reliability of citation counts in reflecting the quality of an academic article depends on the time lapse between the publication date and the timing of observing the number of citations received. Citations observed at a point in time too close to the date of publication will not necessarily offer a quality proxy that is preferable to impact factor. According to Abramo et al. (2010a), if we do not have data on citations counts for at least a period of 2–5 years (depending on the academic field) after the end of the evaluation period considered it will be preferable to use journal impact factor to approximate publication quality. Nevertheless, since the distribution of both citations and journal impact factors are typically skewed to the right in all academic fields it seems appropriate to use the percentile as means of standardization (Abramo et al. 2010a).[17]

When there are data for a sufficient period of time after the end of the evaluation period considered, citations counts can be used not only as a quality ladder to adjust publication counts but also as an additional research output metric accounting for the value of academic achievements. Academic publications embedding new knowledge have different values measured by their impact on academic achievements. Citations represent a proxy measure of the value of research output that is

[15] For example, for a two-author paper, the first author receives 2/3 and the second 1/3 of credit. For a paper where three authors are involved, the first author receives 6/11, the second 3/11 and the third 2/11 of credit.

[16] There is also a disagreement on whether the choice of counting method affects more papers or citations counts. Lin et al. (2013) found that it impacts citation counts more than paper counts while Abramo et al. (2013b) reached the opposite conclusion.

[17] Right-hand skewness implies that most papers are relatively little cited and there are only few papers with many citations, and that the vast majority of papers is published in relatively low impact journals.

usually included in assessment exercises, in spite of limitations due to negative citations and network citations. Note that when counting citations different weights may be given depending upon the citing article influence, the journal in which it is published, etc.

In order to make citation counts a meaningful metric of research value, their total number should be standardized especially for comparisons across fields to reflect differences in citation intensity as well as the various degrees of covering of each academic field in the existing citation databases. This will render citation counts comparable across different research fields and time. Different scaling factors have been proposed in the literature: the arithmetic mean, the geometric mean, the median, the z-score, etc. Due to the (right) skewness of citations' distribution it seems preferable to use the median as a scaling factor. Empirical evidence from bibliometric studies suggests, however, that the arithmetic average seems the most effective scaling factor when the average is based on the publications actually cited and thus excluding those not cited from the calculation of the arithmetic average (i.e. Abramo et al. 2012a, b).[18] Scaling of citation counts is carried out by multiplying the citations of each publication by the chosen scaling factor that characterizes the distribution of citations of articles from the same academic field and the same year.

12.3.3.2 Alternative DEA Models for Assessing Research Productivity

The main purpose of using DEA to assess research productivity is to obtain, through an optimization procedure based on linear programming, *a posteriori* weights to aggregate research inputs and outputs in order to derive a single metric, by means of an efficiency score or a composite indicator, reflecting relative achievement (De Witte and Rogge 2010).[19] The *a posteriori* weights may be variable (i.e., unit-specific) or common, and may or may not reflect (at least partially) experts' or stakeholders' opinions.

The flexible (unit-specific) weights resulting from conventional DEA models reflect its underlying assumption that each evaluated unit is allowed to choose, under certain regulatory conditions, its own set of input and output weights in order to show it in the best possible light relative to other units. It is thus able to exaggerate its own advantages and at the same time to downplay its own weaknesses in order to obtain the maximal possible evaluation score. But if after that it is

[18] Abramo et al. (2012a, b) also provided empirical evidence indicating that rankings of individual researchers obtained under different scaling factors (i.e., average, median, cited papers average, cited papers median) do not show significant discrepancies.

[19] The other two research productivity evaluation methods, namely peer review and bibliometrics, rely respectively on a priori weights reflecting experts or stakeholders opinions or use equal weights and appropriate normalizations/standardization to obtain comparable metrics.

still weaker relative to other units in the sample this cannot be put down to the choice of input and output weights. On the other hand, some authors have argued that comparison and ranking is meaningful only if it is conducted on common grounds and thus favour the use of common but not necessarily equal weights across different outputs. Several variants/modifications, such as common weights DEA and average cross efficiency, have been used for such purposes (e.g. see Oral et al. 2014). Lastly, a combination of *a posteriori* and a priori (i.e., model and experts/stakeholders based) weights may also be possible. Two rather distinct approaches have been used in this respect: peer appraisal by means of cross efficiencies and value judgment DEA. In the former case, the value norms (i.e. DEA weights) of all evaluated units are taken into account when assessing the performance of each unit. In the latter case, the DEA weights assigned to (some or all) inputs and outputs are constrained to satisfy a priori restrictions in order to eliminate the possibility of assigning zero values to particular inputs and/or outputs and more generally to ensure DEA weights accord with intuition.

The second methodological aspect that has to be considered at the outset of the evaluation process is whether resources related to research activities will be taken into account or not. This refers to the choice between measuring efficiency or effectiveness. The former compares the outcome(s) of the research related activities relative to the resources employed for this purpose while the latter compares only the outcome(s) of the research related activities and not the means to achieve them. Conventional DEA models may be used to measure efficiency of research activities while measuring effectiveness is equivalent to the construction of composite performance indicators, which can be done using either DEA-based models such as the benefit-of-the-doubt, (BoD), (e.g. Cherchye et al., 2007) or linear programming models (e.g., Kao and Hung 2003).

The third methodological aspect that has to be considered at the outset of the evaluation process is related to the aggregation level at which the assessment exercise will be conducted. This aggregation level runs from individual level to different degrees of institution/organization aggregation, namely departments, schools/colleges, and the university as a whole. We can thus evaluate research productivity of faculty members as well as of the departments or the universities they belong to. According to Abramo and D'Angelo (2014), for any ranking concerning units that are non-homogenous in their research fields it is necessary to start from the measurement of research productivity at the individual (i.e., faculty members) units and then find an appropriate way to aggregate them. This requires a consistent way to aggregate efficiency and effectiveness scores from the individual to the institution level.

12.3.3.3 Measuring Efficiency and Effectiveness of Research Activities

Output orientation is appropriate for measuring research efficiency since in general the overall objective is not to reduce the input while maintaining constant production but to attempt to maximize production with the resources available.

On the other hand, effectiveness of research activity can be estimated by means of two seemingly similar models that share a common feature: they account only for the output side and thus acting as output aggregator functions. In the input side they rely on Koopman's idea of a person who has at his/her disposal a unitary quantity of an aggregated input that is used for research activities. These two models are the BoD and the Kao and Hung (2003) (K&H) model, which gained increasing popularity in recent years as models used to construct composite indicators. Based on Karagiannis and Paschalidou (2014) we next present and contrast these two models under three different specifications of output weights: variable, common, and restricted.

The BoD model is essentially a tool for aggregating linearly quantitative performance sub-indicators into a single composite indicator when the exact weights are not known a priori (Cherchye et al. 2007). For each evaluated unit, it does so by implicitly assigning less (more) weight to those sub-indicators or performance aspects that the assessed unit is a relatively weak (strong) compared to all other units in the sample. Moreover, the estimated weights are allowed to vary across units and time.

In technical terms, the BoD is a benchmarking model that has a DEA-type structure in the sense that the composite indicator is defined by the ratio of actual to benchmark performance, both of which are given by the weighted sum of the sub-indicators considered. Since the composite indicator is designed to take values in the [0,1] interval, benchmark performance attains by construction the maximum value of one (Cherchye et al. 2007). In determining actual overall performance, the weights are selected in such a way as to maximize the value of the composite indicator of the evaluated unit. This in turn guarantees that any other weighting scheme would worsen the ranking of this unit. Moreover, when these weights are used by any other unit in the sample would not result in a composite indicator greater than one. The resulting weights are determined endogenously by solving for each evaluated unit problem (12.11).

$$
\begin{aligned}
& I^k = \max_{w_i^k} \sum_{i=1}^{N} w_i^k I_i^k \\
& s.t. \sum_{i=1}^{N} w_i^k I_i^j \leq 1, \qquad j = 1, \ldots, K \\
& w_i^k \geq 0, \qquad\qquad\quad i = 1, \ldots, N
\end{aligned}
\tag{12.11}
$$

where $\mathrm{I}_{\mathrm{i}}^{\mathrm{k}}$ is the ith sub-indicator of the kth unit, the higher the value the better, $\mathrm{w}_{\mathrm{i}}^{\mathrm{k}}$ are the weights to be estimated, j is used to index units and i to index sub-indicators which in our case correspond to different research outputs (i.e., types of publications, citations, patents).

The BoD model is equivalent to the multiplier form of the Charnes et al. (1978) input-oriented, constant returns to scale (CRS) DEA model when there is a single constant input that takes the value of one for all evaluated units.[20] Based on this, the dual formulation of the BoD model is given as (12.12).

$$
\begin{aligned}
& I^k = \min_{\lambda_j^k} \sum_{j=1}^{K} \lambda_j^k \\
& s.t. \sum_{j=1}^{K} \lambda_j^k I_i^j \geq I_i^k, \qquad i = 1, \ldots, N \\
& \lambda_j^k \geq 0, \qquad\qquad\quad j = 1, \ldots, K
\end{aligned}
\tag{12.12}
$$

where λ refers to intensity variables. This implies that the value of the composite indicator is in fact equal to the sum of the intensity variables. From the inequality constraints on the intensity variables it is clear that the BoD model exhibits constant returns to scale.

On the other hand, the Kao and Hung (2003) model has a similar structure in the sense of deriving a set of *a posteriori* weights that maximize the value of a composite research performance indicator but now under the assumption that this set of weights satisfies for each evaluated unit an adding-up/normalization constraint. The K&H model is written as (12.13):

$$
\begin{aligned}
& E^k = \max_{u_i^k} \sum_{i=1}^{N} u_i^k I_i^k \\
& s.t. \sum_{i=1}^{N} u_i^k = 1 \\
& u_i^k \geq 0, \qquad\qquad i = 1, \ldots, N
\end{aligned}
\tag{12.13}
$$

Even though the two models have the same objective function they differ in terms of the underlying constraints, which in the case of the K&H model render a linear programming rather than a DEA-type model. In the K&H model there is only one (equality) constraint, besides the non-negativity constraints of the weights, while in the BoD model the number of (inequality) constraints is equal to the number of evaluated units.

Besides these differences, Kao et al. (2008) have shown that the two models are related to each other as long as the set of sub-indicators to be aggregated are normalized at the outset to be within the [0,1] interval; that is, $0 \leq \mathrm{I}_\mathrm{i}^\mathrm{k} \leq 1 \ \ \forall\, \mathrm{i} = 1, \ldots, \mathrm{N}$. In this case one can verify that $\mathrm{E}^\mathrm{k} = \mathrm{I}^\mathrm{k}/\mathrm{W}^\mathrm{k}$ where $\mathrm{W}^\mathrm{k} = \sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{w}_\mathrm{i}^\mathrm{k}$ and $\mathrm{u}_\mathrm{i}^\mathrm{k} = \mathrm{w}_\mathrm{i}^\mathrm{k}/\mathrm{W}$. This implies that the K&H model delivers values of the composite indicator that are close but

[20] More on the radial DEA models with a single constant input can be found in Lovell and Pastor (1999), Caporaletti et al. (1999) and Liu et al. (2011). Notice also that unitary input DEA models are equivalent to DEA models without explicit inputs.

not always equal to those suggested by the BoD model. More importantly, Karagiannis and Paschalidou (2014) note that while from the BoD weights we can derive the weights implied by the K&H model the converse is not possible. This limitation of the K&H model is related to the types and number of constraints that it involves. On the other hand, for this same reason, the K&H model is computationally less demanding. Lastly, at present the K&H model has unknown aggregation properties and thus we cannot move the analysis of research productivity from the individual to institution (i.e., department or university) level in a theoretically consistent way.

In contrast, such an aggregation rule for the BoD model has been developed by Karagiannis (2016) within the framework of aggregate efficiency scores. In particular, it has been shown that the arithmetic average (in (12.14)) is the theoretically consistent aggregation rule for the BoD model.

$$I = \frac{1}{K}\sum\nolimits_k I^k \tag{12.14}$$

Thus, the aggregate composite performance indicator equals the simple (un-weighted) arithmetic average of the estimated individual composite indicators. This results from the single constant (unitary) input structure of the BoD model and the denominator rule (Färe and Karagiannis 2013) stating that consistency in aggregation of ratio-type performance measures, including efficiency indices, is ensured as long as the weights are defined in terms of the variable being in the denominator.[21] For an input-oriented model such as the BoD, these will be actual cost or input shares. But since all evaluated units have the same amount of (one unit) and face the same price for the single input, the share weights become equal to 1/K. In terms of research activity, this result implies that a department's research productivity can be simply estimated by means of the average research productivity of its faculty members.

Regarding now the estimation of research effectiveness in terms of common instead of variable weights, which according to Kao and Hung (2005) and Wang et al. (2011) among others have the advantage of making it possible to compare and rank the performance of all evaluated units and not only classify them as efficient or inefficient, both the BoD and the K&H models possess special features.[22] First, for the BoD model Karagiannis and Paleologou (2014) have shown that common weights are related to average cross efficiency, which is of particular interest in assessing research productivity (Oral et al. 2014) as it provides the basis of giving the right to every faculty member to have a "say" about the performance of other faculty members in the same institution. In particular, average cross efficiency in the BoD model is based on a set of common weights given by the simple arithmetic average of weights obtained from the self-appraisal version of the model, i.e., the

[21] By consistency here we mean that the resulting aggregate measure has exactly the same intuitive interpretation as the individual efficiency scores.

[22] Another advantage of common weights is that they can be applied to calculate performance indices for DMUs not in the sample (Kao and Hung 2007).

one discussed above. On the other hand, a common set of weights in the K&H model can be obtained by applying Kao and Hung (2005) compromise solution. That is by running a linear ordinary least squares regression (not including an intercept term) of the composite indicator obtained from the conventional form of the model, on the set of sub-indicators under the restriction that the estimated parameters sum up to one.

Finally, a set of weights on which a research productivity assessment may be carried out is that reflecting value judgment. For the BoD model, this is incorporated in terms of weights restrictions in the multiplier form of the model. Several types of weights restrictions have been used for this purpose including pie shares (e.g. Cherchye et al. 2007) and partial descending ordering (i.e., $w_1^k > w_2^k > w_3^k > \ldots$). The latter is a case of particular interest for the K&H model because then as Ng (2007, 2008) has shown there is no need to estimate the composite performance indicator by means of linear programming but rather to compute it based on partial averages; that is, the composite indicator is given as in (12.15), where i is used to index sub-indicators.

$$Max\left\{I_1^k, \frac{\sum_{i=1}^{2} I_i^k}{2}, \frac{\sum_{i=1}^{3} I_i^k}{3}, \ldots\right\} \tag{12.15}$$

Lastly, the BoD model, as a DEA-type model, can also be used to examine research productivity over time by means of the corresponding technology-based (i.e. Malmquist or Hicks-Moorsteen) indices, an aspect of performance evaluation that cannot be done with the K&H model. For the BoD model, Karagiannis and Lovell (2016) have shown that *first,* the Malmquist and Hicks-Moorsteen productivity indices coincide, they are multiplicatively complete,[23] and the choice of orientation for the measurement of productivity change does not matter. *Second,* there is a unique decomposition of the sources of productivity change containing three independent components, namely technical efficiency change, neutral technical change and output biased technical change. *Third,* the aggregate output-oriented Malmquist productivity index is given by the geometric average between any two periods of the simple (un-weighted) arithmetic average of the individual contemporaneous and mixed period efficiencies.

12.3.3.4 An Illustrative Application

The empirical application is from Karagiannis (2016) who applied the BoD model to evaluate the research achievements of faculty members in the Department of Economics at the University of Macedonia, Greece during the period 2000–2006. In

[23] A productivity index is multiplicatively complete if it can be written in a ratio form of input/output indices that are non-negative, non-decreasing, linearly homogenous scalar functions (O'Donnell 2012).

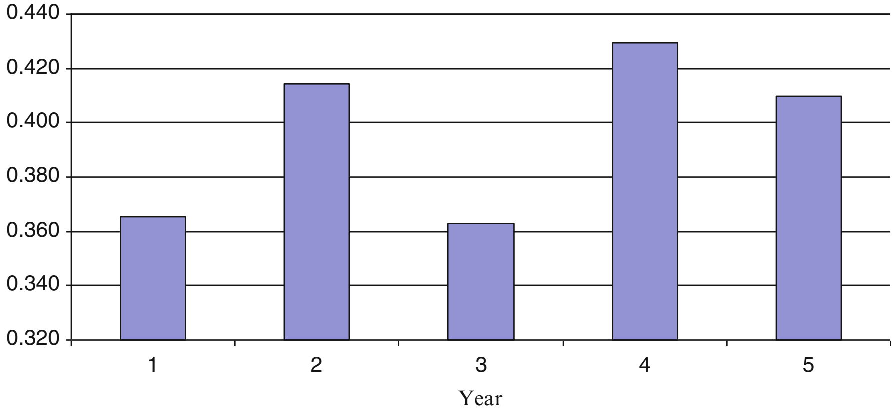


Fig. 12.9 Average annual technical efficiency: Source Karagiannis (2016)

the proposed setting the single constant input corresponds to each faculty member and two outputs were considered, namely, journal articles and all other publications, which are measured by whole numbers. As journal articles are considered all publications in outlets referenced in the Journal of Economic Literature and as other publications are considered papers published in journals not referenced in the Journal of Economic Literature, chapters in books and edited volumes. The relevant data reveal that, on average, each faculty member published almost one journal paper per year during the period 2000–2003 and there seems to be an improvement in research achievements as the annual average increases to somewhat above one during the period 2004–2006. The corresponding figures for other publications are well below one for the whole period, with a trend to decline significantly in the last 2 years. In addition, both kinds of publications are unevenly distributed between faculty members. There are a few faculty members with satisfactory achievements in journal article publications (more than two and a half on average per year) and one faculty member with similar performance in terms of other publications but most of them are around the departmental average. There were however two faculty members that had no journal article published during the whole period under consideration and with only one other publication each. Moreover, the achievements of newcomers are underestimated because of no entries in the data over the whole period under consideration. For convenience these two cases were disregarded and thus a total of 20 faculty members are included in the sample.

The average annual scores of technical efficiency (see Fig. 12.9), which reflect research efficiency at the department level, were found to be rather low and in the range of 0.36–0.43 indicating the relative heterogeneity in the achievements of faculty members.

12.3.4 Using DEA to Assess Administrative Services in Universities

The bulk of applications of DEA in Higher Education have focused either on assessments from a broadly academic perspective in terms of effectiveness of value added in teaching and research or in terms of economic efficiency at institution or department level. One DEA application which has addressed efficiency at university function level is that by Casu and Thanassoulis (2006) which has looked into university administrative services in the UK.

Universities as all publicly funded services are tasked to achieve value for money. Typically the focus when looking at universities has been on teaching and research. The allocation of resources between academic and non-academic departments has rarely, if ever, been subject to scrutiny. Yet, administrative expenditure is substantial. For example, in 1997/1998 expenditure on Administrative and Central Services (excluding services such as premises and catering) represented some 12 % of total UK higher education sector and nearly 30 % of expenditure on academic departments (Casu and Thanassoulis 2006). Using data for 1999/00, (Casu and Thanassoulis 2006) assessed expenditure on central administrative services (CAS) in UK universities, in order to identify the scope for potential savings in this area. A follow up study, as we will see later, used longitudinal data and the Malmquist index to measure the change in productivity in administrative services in UK Universities over the period 1999/00 to 2004/5.

University administrative services are organised at varying levels of devolvement to departments. The scope of functions (e.g. finance, personnel etc.) to be included within delineated units of assessment was decided with the aid of an advisory board of senior university academics and administrators (Casu et al. 2005). A follow up workshop was organised involving individuals at all levels of university administration, both academic and non-academic. A computer-mediated Group Support System (GSS) was used to home in on possible input-output variables to use in a DEA framework. The use of computer mediation was deemed beneficial by, for example, removing common communication barriers such as being interrupted, dominating discussants or a reluctance to share views. The software used was ('JOURNEY' – JOintly Understand, Reflect and Negotiate) to structure the discussion. Details of the facilitation leading to the delineation of the CAS unit of assessment and the related input-output variables can be found in (Casu et al. 2005). We summarise here the conclusions reached.

The services within a broad definition of CAS are illustrated in Fig. 12.10.

However, it was broadly agreed that in most universities services such as library, catering, residences etc. are self-contained both in organisational terms and in terms of data available. Hence they can be assessed separately. Thus the authors excluded self-contained functions to define CAS as the Core component in Fig. 12.10.

The main stakeholders were identified as:

– Students,
– Non-administrative Staff,

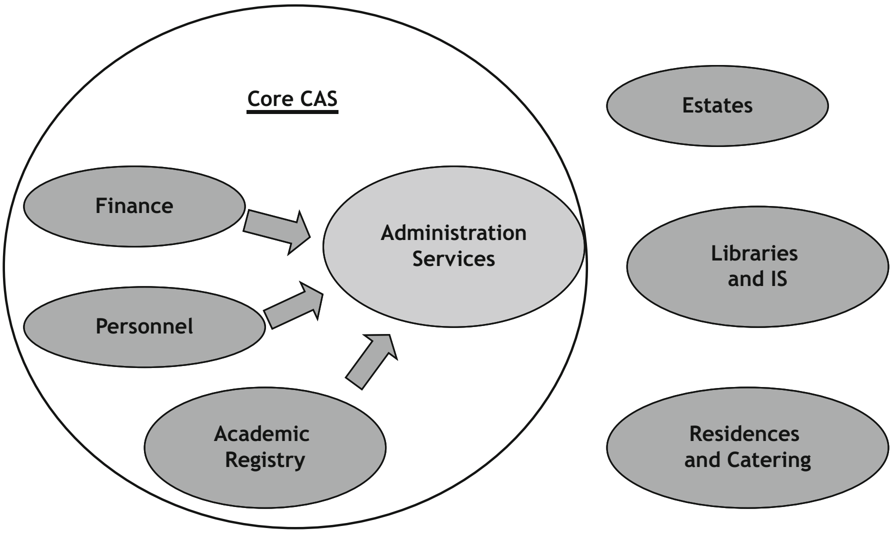


Fig. 12.10 Services included in CAS

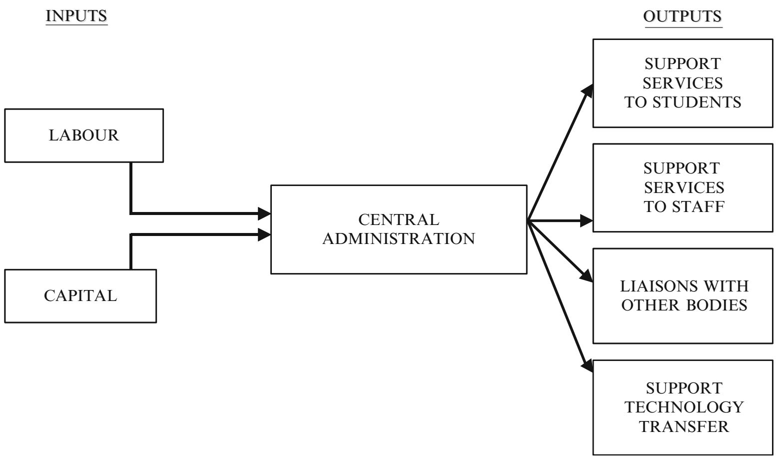


Fig. 12.11 A conceptual framework of the CAS unit of assessment

– Suppliers,
– Funders,
– Community,
– Policy Makers and
– Educational and Industrial Partners.

This led to the conceptual model for identifying the input-output variables for CAS in Fig. 12.11 (Casu and Thanassoulis 2006).

Table 12.4 Inputs and outputs used in Casu and Thanassoulis (2006)

Inputs	Outputs
Total administrative costs (total administrative staff costs + other operating expenses)	Total income from students
	Total staff costs (minus total administrative staff costs)
	Technology transfer

Technology Transfer covers such activities as consultancy, joint ventures between the university and external organisations, setting up commercial companies to exploit the results of research and so on

Table 12.5 Modified input and output variables

Administrative cost measures (inputs)	Activity volume measures (outputs)
Administrative staff costs	Admin services for students Measure: *total income from students*
	Admin services for non-admin staff Measure: *non-administrative staff cost*
Operating expenditure on administration excluding staff costs	Services for research grants and other services rendered Measure: *income from these sources- using 3 year moving average*
	Services for research volume and quality Measure: *QR component of HEFCE income*

Secondary data as returned by Universities to the Higher Education Statistics Agency (HESA) for the year 1999/00 were used to operationalise the above framework for DEA purposes into the input-output variables in Table 12.4.

A total of 108 university administrations were assessed (England (86), Wales (7), Scotland (13) and Northern Ireland (2)). The assessment was run as an input minimisation, variable returns to scale DEA model. Considerable inefficiency was found – mean level of inefficiency 26.6 % – suggesting about a quarter of administrative expenditure in the universities assessed can be saved. This varied of course across institutions. Some 17 of them were identified as 'efficient' in the sense that no scope for savings could be identified in CAS services relative to other institutions.

The findings of the assessment need to be seen as indicative rather than definitive as the authors did not have detailed enough data as to what part of non academic staff costs in academic departments relates to the administration functions modelled. Moreover, the data returned to HESA by universities may not be fully comparable as institutions have some latitude in interpreting the data definitions.

The foregoing assessment was followed by an unpublished assessment of the change in productivity of administrative services of UK Universities between 1999/00 and 2004/5. (*The description here is drawn from the report submitted to the funders of the project*). Following the initial assessment using the 1999/00 data feedback from universities (a consultation event was attended by over 100 representatives from UK University administrations) the original input-output set was modified to the set shown in Table 12.5.

There are two main changes from the initial assessment; one is that two input variables are used, and the other is that a fourth output was added, (*QR component of research income*). One other change was that non administrative staff costs were adjusted to account for clinical staff present in some universities who have a substantially higher salary than non clinical staff.

The separation of Administrative Staff costs from Operating Expenditure (OPEX) was because experts could not agree whether the two types of resource can substitute for each other. The balance of view was mostly that they cannot substitute for each other, but data was not available to net out items of OPEX such as IT which can substitute for staff. There was a further pragmatic reason, apart from the issue of substitutability that led to the decision to focus on modelling the two inputs separately. Staff costs in comparison to OPEX are both more homogeneous across institutions and more clearly attributed to central and/or academic department administration. Modelling staff costs separately means that one can avoid contaminating this more clearly identifiable expenditure with the more heterogeneous and not easily attributable OPEX. This in turn made it possible to arrive at results which for staff expenditure would be much more reliable. In the event the authors ran three models, one using each input separately and a third using the two inputs jointly.

The fourth output added, *Quality-Related* (QR) research income, refers to the component of funding the institution receives for research as distinct from ad hoc research grants academic staff may secure through bidding. The QR component is based on the research quality rating achieved by each academic cost centre of that university in what at the time was the **research assessment exercise** (RAE). The amount also reflects the number of research active staff submitted by the University to the RAE. The QR output was intended to capture the administrative effort that academic research imposes on administrators over and above that already reflected in the other three outputs in the model. A weight restriction was used in the DEA model to ensure that one unit of QR funding was not deemed by the model to require more administrative (input) resource than an equal monetary unit relating to any one of the remaining three outputs. The QR component was also deflated to account for the fact that some disciplines receive higher funding than other disciplines for the same research quality rating depending on whether research requires laboratories etc.

Malmquist indices of productivity change were computed using the above input-output set with data for the period 1999/0-2004/5. The findings in summary were as follows:

- The scope for efficiency savings in administrative staff costs is of the order of 25 % in 1999/00 on average and it drops to about 20 % by 2004/5. There are some 20 units with good benchmark performance without any identified scope for efficiency savings. About half of these benchmark units are found to be consistently cost-efficient in all 6 years of our analysis. These units could prove examples of good operating practices in administration, which need to be identified by a careful study of the units beyond the data returned to HESA that were used.

- The picture in terms of OPEX efficiency is similar. The scope for efficiency savings is somewhat higher, between 20 % and 25 % on average each year. Again, some 20 units are benchmark and most of them are consistently so over time, offering the prospect of good operating practices.
- When taking administrative staff and OPEX as substitutable and using them jointly as inputs it is found that the scope for efficiency savings drops to about 15 % per annum in each one of the two inputs. The reduced scope found is because now benchmark units must offer low levels both on staff cost and OPEX rather than just on one of the two. As with the preceding two assessments here too it is found that a considerable number of units are benchmark in all 6 years which therefore could prove exemplars for other units to emulate. At the other extreme, some of the units with large scope for efficiency savings are so on all the 6 years suggesting persistent issues of expenditure control.
- Looking at productivity change between 1999/00 and 2004/5 a divergent picture is found between administrative single and two-input models. In the case of administrative staff taken as a self-contained resource it is found that there is on average a drop in productivity so that for given levels of the proxy output variables staff cost is about 95 % in 1999/00 compared to what it is in 2004/5, at constant prices. Looking deeper it is found that generally units do keep up with the benchmark units but it is the benchmark units that are performing less productively by almost 7 % in 2004/5 compared to 1999/00. The situation is not helped by a slight loss of productivity through scale sizes becoming less productive, by about 1.5 % in 2004/5 compared to 1999/00.
- In contrast when we look at OPEX as a self-contained resource or indeed at administrative staff and OPEX as joint resources it is found that productivity is more or less stable between 1999/00 and 2004/5. There is a slight loss of about 2 % but given the noise in the data this is not significant. What is significantly different between administrative staff on the one hand and OPEX or the two joint inputs on the other is that benchmark performance improves on average by about 8 % between 1999/00 and 2004/5. That is for given proxy output levels OPEX costs or joint OPEX and staff costs drop by about 8 % in 2004/5 compared to 1999/00. Unfortunately non benchmark units cannot quite keep up with this higher productivity. Also there is a slight deterioration again in scale size efficiency and the net effect is that despite the improved benchmark productivity, productivity on average is stable to slightly down between 1999/00 and 2004/5.
- As with cost efficient practices here too the analysis identified a small set of units which register significant productivity gains and others which register significant productivity losses between 1999/00 and 2004/5. An investigation of the causes in each case would be instrumental for offering advice to other units on how to gain in productivity over time and avoid losses. Such investigations need access to the units concerned beyond the HESA data used in the analysis.

12.3.5 Using DEA to Rank Universities

Universities are ranked in numerous ways, as a whole or by department by the popular press (De Witte and Hudrlikova 2013).[24] For example in the UK the Times, the Sunday Times and The Guardian are just some of the papers publishing so called League Tables. However, as universities are multi outcome entities the issue arises how to weight individual indicators of university performance such as research outcomes, teaching quality, employability of graduates etc. The DEA-based BoD model outlined earlier in Sect. 12.3.3 can be used to determine an endogenous weighting system, permitting each university to give weights that show it in the best light relative to other institutions. The rationale for this is that institutions should be allowed to have their own areas of excellence the ranking should be able to reflect this (e.g. see van Vught and Westerheijden 2010).

Let I_i^k be the index reflecting the position of university k on outcome i relative to other universities and let there be n outcomes on which universities are to be compared. It is assumed that the larger the value of I_i^k the better university k is relative to other universities. The BoD score I^k of university k can be computed through model (12.11) in Sect. 12.3.3. The BoD model chooses the weights w_i^k for each university k which maximize its ranking score I^k. The evaluated university 'can' choose the weights to maximize its 'aggregate' I^k. The weights are restricted not to permit any other university using those same weights to have a I^k value above 1. Because of this constraint the best performing universities will obtain a BoD score equal to 1. The rest of the universities will obtain a BoD score I^k lower than 1. The difference $(1 - I^k)$ expresses the shortfall in institutional attainments relative to other universities. The BoD scores can be used to rank universities.

De Witte and Hudrlikova (2013) deploy the foregoing method to rank the 200 universities originally ranked by The Times Higher Education Supplement (THES) in 2009. They use the variant outlined above which included repeated sampling with replacement and recomputation of ranks in the framework of the Cazals et al. (2002) approach. This mitigates the impact of outlying observations which might arise from measurement errors or from atypical observations (e.g., due to historic decisions). They use the same variables to rank universities as THES. These are:

- the reputation of the university (THES weight of 40 %)
- opinion on the university by employers (THES weight of 10 %)
- research excellence (THES weight of 20 %)
- staff-to-student ratio (THES weight of 20 %)
- international faculty (THES weight of 5 %)
- overseas students (THES weight of 5 %)

[24] This section is based on De Witte and Hudrlikova (2013).

They also control for contextual variables which might advantage or disadvantage some institutions. These are

– tuition fees,
– size of the university,
– research output (relative to size and faculty areas) and
– origin in an English speaking country.
– university autonomy.

The approach used to control for the foregoing contextual variables is that of De Witte and Kortelainen (2013). It was found that when universities are assessed accounting for exogenous background characteristics, European universities take the top 15 places. The original top universities according to the (unconditional) THES move from place 1 (Harvard University) to place 18, for Cambridge from 2 to 21 and Yale University from 3 to 24. These results suggest that it is important to control for contextual factors in comparing universities (see De Witte and Hudrlikova 2013 for an extensive discussion). Moreover, the results show that giving institutions the 'benefit of the doubt' on the multiple dimensions of institutions is important. As the nature of institutions differs, league tables should give credit to the relative strengths of each institution. DEA is a promising methodology for doing so.

12.3.6 Network DEA

All of the DEA models considered above are 'black box' in nature – that is, they consider the production process as one of conversion of inputs into outputs without further consideration of the mechanism by which this is done. An alternative approach, pioneered by Färe (1991) and Färe and Grosskopf (1996b), and more recently refined by Tone and Tsutsui (2009) involves constructing a network of nodes within the production unit; these nodes each have inputs and outputs, with some outputs of some nodes serving as inputs into other nodes. The overall efficiency of the production unit as a whole is then a function of the efficiency with which each node performs its role.

A network DEA of this type has been used to assess English higher education institutions by Johnes (2013). The network is illustrated in Fig. 12.12. The model shows two nodes. Node 1 uses measures of intake quality, student-staff ratio and per student spend to deliver degree results as an intermediate output. The degree results and the research reputation of the institution are inputs to Node 2 from which the ultimate outputs of the system are employability and student satisfaction. The results indicate that, while institutions are typically highly efficient in converting inputs into two of the outputs – student satisfaction and degree results – their performance in converting inputs into employability is less impressive, with only four institutions scoring above 90 % at the second node. This suggests that institutions looking to improve their overall efficiency might find quick wins in this area.

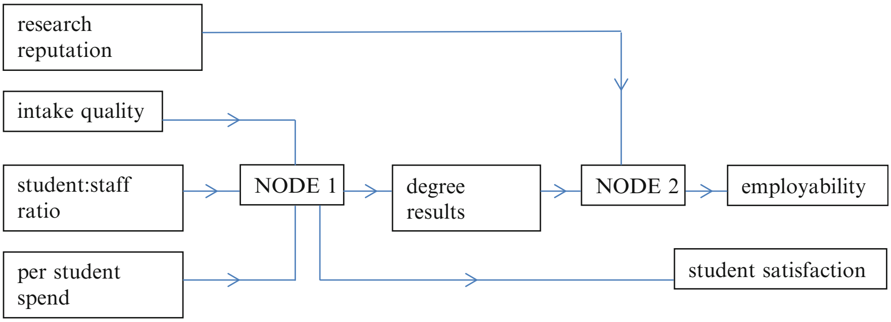


Fig. 12.12 A network DEA model

That said, the network structure used in this analysis is imposed by the analyst, and may be contentious. Different network designs are likely to lead to different results. Much work therefore remains to be done in the application of network models in the sphere of higher education.

12.4 Applications of DEA in Higher Education: Person Level

The foregoing DEA applications have focused on efficiency assessments broadly around the perspective of academic value added by a secondary school or a higher education institution. In this section we look at the use of DEA in assessing higher education academic staff on their key functions: research and teaching. De Witte et al. (2013a, b) discuss in a DEA model how the two activities relate, and whether there are economies of scope between teaching and research. This section focusses on teaching and research as two separate activities.

12.4.1 Assessing Academics on Research Output

Universities and colleges are increasingly interested in evaluating the performance of their academic staff, both in terms of teaching and research.[25] Current literature on research evaluation mainly employs single-criterion measures, such as reputational ratings gathered by peer reviews, number of publications in a predefined set of refereed journals, or citation counts (see De Witte and Rogge 2010). Recently, several authors have criticized such simplistic measures doubting whether they are able to accurately convey research performance. They argue the nature of research

[25] This section is based on De Witte and Rogge (2010).

is far too complex to be reflected in a single measure. A score to measure research should be multidimensional, control for exogenous characteristics and account for the different preferences by the different stakeholders and individuals. Therefore, the construction of a multi-criteria Research Evaluation Score (RES-score) is an intricate matter with, amongst others, two important conceptual and methodological difficulties to overcome:

1. How should one weight and aggregate the different research output criteria? Or, stated differently, how important are the several research outputs in the overall performance evaluation? Is it legitimate to assign a uniform set of weights over the several research output criteria (i.e., equal/fixed weights)? Also, is it legitimate to apply a uniform set of weights to all evaluated researchers? Some researchers may focus on writing journal papers or books, while others in attracting research funding. Using the same weights for all researchers, could be seen as unfair within a research unit.
2. How should the RES-scores be adjusted for the impact of exogenous characteristics which are (often) beyond the control of the researcher? There are numerous findings in the academic literature which suggest that some background characteristics (e.g., age, gender, rank/tenure, time spent on teaching, department policy, etc.) may have a significant impact on the research performance of academic staff. Yet, traditional RES-scores do not account for differences in these uncontrollable conditions. Consequently, these scores are inherently biased towards researchers working under more favourable conditions.

The few studies which use multi-criteria instruments, calculate commonly a *global* RES-score for an individual as an arithmetic mean or a weighted sum of the researchers' performance on several criteria. They compute for researcher k the score as shown:

$$RES_k = \sum\nolimits_{i=1}^{N} w_i^k I_i^k$$
$$\sum\nolimits_{i=1}^{N} w_i^k = 1 \tag{12.16}$$

where I_i^k is the number of outputs researcher k has in category i; w_i^k is the weight assigned to output category i for researcher k (with $0 \le w_i^k \le 1$ and $\sum\nolimits_{i=1}^{N} w_i^k = 1$); N is the number of output categories considered in the research evaluation. In studies where the RES-scores are computed as an arithmetic mean we have $w_i^k = 1/\mathrm{N}$. This implies that all aspects of research output are assumed to be of equal importance. In essence, an arithmetic mean RES-score corresponds to a measure where the publications are just counted over the different research output categories without any correction for their quality. When the RES-score is constructed as a weighted sum of publications with w_i^k varying over the different research output categories, this score corresponds essentially to a simple publication count with a correction for quality.

To account for the weighting issues in the construction of the research evaluation scores, De Witte and Rogge (2010) propose a specially tailored version of the '**benefit-of-the-doubt**' model. This data-driven weighting procedure has five important advantages compared to the traditional model as in Eq. (12.16).

First, for each evaluated researcher, weights for the various output criteria are chosen such that the most favourable RES-score is realized. One could intuitively argue that, given the uncertainty and lack of consensus on the true weights of research outputs, BoD looks for those weights w_i^k which put the evaluated researcher k in the best possible light compared to his/her colleagues. As such, the research performance measure is relative. The BoD model grants the 'benefit-of-the-doubt' to each researcher in an already sensitive evaluation environment. Being evaluated optimally, disappointed researchers (i.e., researchers with RES-scores below expectations) cannot blame these poor evaluations to subjective or unfair weights. Any other weighting scheme than the one specified by the BoD model would worsen their RES-score relative to that of others. Second, the BoD model is flexible to incorporate stakeholder opinion (e.g., researchers, faculty administrators, experts) in the construction of the RES-scores through pre-specified weight restrictions, to ensure that importance values are chosen in line with 'agreed judgments' of these stakeholders, without pre-determining the exact weights. Third, researchers are evaluated relative to the observed performances of colleagues. This clearly marks a deviation from the common practice in which benchmarks are exogenously determined by department administrators often without any sound foundation. Fourth, we can adjust the BoD model such that its outcomes are less sensitive to influences of outlying or extreme observations as well as potential measurement error in the data, e.g. by using the robust order-*m* method of Cazals et al. (2002), adapting it to the BoD setting. Finally, the BoD model can be adjusted (after the conditional efficiency approach of Daraio and Simar (2005, 2007a, b)) to account for background influences (e.g., age, gender, rank, PhD, teaching load, time for research, etc.).

De Witte and Rogge (2010) applied the BoD approach on a dataset collected at the Department of 'Business Administration' of the Hogeschool Universiteit Brussel (Belgium) in the academic years 2006–2007 and 2007–2008. They argue their approach enables one to include different (potentially) influential conditions (outside the control of the researcher) into the built-up of the overall RES-scores. Hogeschool Universiteit Brussel (HUB) resembles in many ways 'new' (former polytechnic) universities in the UK and the colleges in the US. In particular, it used to be an educational institution with exclusive focus on teaching, but recently, thanks to the Bologna reforms (and a research focus process initiated by the Government of the Flemish part of Belgium), Hogeschool Universiteit Brussel became increasingly research-oriented. The data set comprised research output data on all 81 research staff of the University. They added to this data on age, gender, doctoral degree, tenure, (official) teaching load, and (official) time for research. The data were further enriched with a questionnaire on the researcher's opinions and perceptions on research satisfaction and personal goals.

The core BoD idea is that output criteria on which the evaluated researcher performs well compared to his/her colleagues in the reference set Υ, should weight more heavily than the output criteria on which he performs relatively poor. The rationale for doing so is that a good (poor) relative performance is considered to be an indication of a high (low) importance the evaluated researchers attaches to each criterion. For example, if, in comparison to his/her colleagues, the researcher under evaluation published a high number of papers in international journals this reveals that the researcher considers such publications to be of high importance. Consequently, his/her performances should weigh more heavily this criterion (i.e., high weight w_i^k). In other words, for each researcher separately, BoD looks for the weights that maximize (minimize) the impact of the criteria where the researcher performs relatively well (poorly) compared to other researchers. Hence, BoD-weights w_i^k in this sense are optimal and yield the maximal RES-score to the individual concerned (see (12.11) in Sect. 12.3.3).[26]

The BoD model for researcher k lets the data speak for themselves and endogenously selects those weights w_i^k which maximize his/her RES-score. Any other weighting scheme than the one specified by the BoD model would worsen the indicator RES_k for researcher k. This data-orientation is justifiable in the context of evaluating research performance where there is usually a lack of agreement among stakeholders (i.e., policy makers, researchers, etc.), and uncertainty about the proper importance values of the research output criteria. This perspective clearly deviates from the current practice of using single-criterion measures or multiple-criteria as in (12.16) with or without a correction for the perceived quality. By allowing for 'personalized' and 'optimal' weight restrictions, the BoD model is clearly more attractive to the individual researchers. To a certain extent (i.e., the weight bounds), researchers are given some leeway in their publication outlets. As such, the BoD model is less restrictive than a RES score based on pre-determined weights.

Note that the standard BoD model as in (12.11) grants evaluated researchers considerable leeway in the choice of their most favourable weights w_i^k. Only two constraints have to be satisfied. The first one is the 'normalization' constraint that ensures that all RES-scores computed with the evaluated researcher's most favourable weights w_i^k, can at most be unity (or, equivalently, 100 %). Thus, we obtain $0 \leq RES_j \leq 1$ (j = 1,. . .,k,. . .,N) with higher values indicating better overall relative research performance. The second is the set which limits weights to be non-negative ($w_i^k \geq 0$). Apart from these restrictions weights can be chosen completely free to maximize the RES-score of the evaluated researcher vis-à-vis those of other researchers. However, in some situations, it can allow a researcher to appear as a brilliant performer in a way that is difficult to justify. For instance, while having no publications in any but one research output criterion, which may be generally not highly regarded, a researcher could place a high weight on that

[26] For completeness, we mention that BoD alternatively allows for a 'worst-case' perspective in which entities receive their worst set of weights, hence, high (low) weights on performance indicators on which they perform relative weak (strong) (Zhou et al. 2007).

criterion and zero weights on all other criteria and achieve a high RES-score without violating the model restrictions. In such research evaluations, RES-scores reflect the researchers' performance on one single dimension. More generally it is possible for the BoD model to yield weights which deviate too much from what stakeholders (i.e., the faculty board, evaluated academics) would believe is appropriate. Without doubt, opponents of research evaluations will claim that RES-scores based on improper weights are not meaningful.

BoD models can be modified to incorporate weights restrictions as in traditional DEA models. Formally, this involves adding the general weight constraint (c) to the standard BoD-model:

$$w_i^k \in \mathrm{W_e} \quad \mathrm{i} = 1, \ldots, \mathrm{q} \text{ and } \mathrm{e} \in \mathrm{E} \tag{c}$$

with $\mathrm{W_e}$ denoting the set of permissible weight values defined based upon the opinion of selected stakeholders $e \in E$. It is crucial for the credibility and acceptance of RES-scores to define weight restrictions which reflect stakeholder opinions when available.

Using $w_i^k = \mathrm{w_{k,i}}$ De Witte and Rogge (2010) used the ordinal ranking of nine research output criteria, as agreed by stakeholders as in (12.17).

$$\mathrm{w_{k,1}} = \mathrm{w_{k,2}} \geq \mathrm{w_{k,3}} = \mathrm{w_{k,4}} \geq \mathrm{w_{k,5}} \geq \mathrm{w_{k,6}} = \mathrm{w_{k,7}} \geq \mathrm{w_{k,8}} = \mathrm{w_{k,9}} \geq 0.01 \tag{12.17}$$

From a technical perspective, we have to adjust these additional weight restrictions for the potential presence of zero values in the evaluation data. Indeed, in one or multiple output dimensions researchers may not have been able to produce any publication during the evaluation period (hence, the associated I_i^k's are equal to zero). The endogenous weighting procedure of BoD will automatically assign a zero weight to such output criteria. However, in our evaluation procedure (with the additional ordinal weight restrictions as specified above), this standard procedure may lead to infeasibilities. Kuosmanen (2002) and Cherchye and Kuosmanen (2006) proposed a simple modification of the weight restriction to prevent this infeasibility: multiply the constraints by the product of the corresponding I_i^k's.[27] Formally,

$$\begin{aligned}
&(\mathrm{w_{k,1}} - \mathrm{w_{k,2}}) \times I_1^k \times I_2^k = 0 \\
&(\mathrm{w_{k,2}} - \mathrm{w_{k,3}}) \times I_2^k \times I_3^k \geq 0 \\
&\qquad\qquad \cdots \\
&(\mathrm{w_{k,6}} - \mathrm{w_{k,9}}) \times I_6^k \times I_9^k = 0 \\
&\mathrm{w_{k,i}} \times I_i^k \geq I_i^k \times 0.01 \qquad \forall\, \mathrm{i} = 1, \ldots, 9
\end{aligned} \tag{12.18}$$

[27] See Kuosmanen (2002) for a more comprehensive discussion.

In this adjusted version of the additional weight restrictions, a standard weight $w_{k,i} = 0$ for an output criterion i with $I_i^k = 0$ no longer forces other weights to be zero. In cases where one or both of the associated I_i^k's equal zero, the restriction becomes redundant and hence has no further influence on the other restrictions in (12.14). If none of the associated I_i^k's are zero, then the adjusted version of the weight restriction reduces to the original restriction as in (12.17).

De Witte and Rogge (2011) suggest two further improvements to the BoD model. First, they suggest to use a robust version such that the BoD model accounts for outlying observations without losing information due to removing such observations from the data set. Second, they suggest a conditional robust version. To do so, they apply the methodology suggested by De Witte and Kortelainen (2013), who extended the conditional efficiency framework to include discrete variables. The conditional efficiency framework compares like with like by computing for each observation a reference set with similar features. The BoD model is then estimated on this reference set with only comparable observations.

In a competitive context (e.g., for personnel selection decisions), by comparing researchers, the conditional RES-scores, which account for exogenous characteristics, can be deemed 'fairer' than unconditional RES-scores. Besides the employment conditions as retention, teaching load and research time the model used by the authors accounted for certain researcher background characteristics such as gender, age, PhD and being a guest researcher at University KU Leuven. Thus in effect their model controls for both exogenous factors (such as age, gender) and for factors which are exogenous to the researcher but not to the University (e.g. a university decision (e.g., hiring faculty without PhD, retention). The latter group of variables is interesting as it is at the discretion of the university. Although this set of background variables is not exhaustive, it contains the variables that the faculty board at HUB (i.e., a mixture of policy makers and researchers) consider as appropriate. Accounting for background variables, the conditional RES estimates increase dramatically. A larger group of researchers (75 %) becomes significant while the median researcher can improve her/his research performance by 21 % (see De Witte and Rogge 2010 for an extensive discussion).

12.4.2 Assessing Academics on Teaching

Students' evaluations of teaching (SETs hereafter) are increasingly used in higher education to evaluate teaching performance.[28] Yet, for all their use, SETs continue to be a controversial topic with teachers, practitioners, and researchers sharing the concern that SET scores tend to be 'unfair' as they fail to properly account for the impact of factors outside the teacher's control (De Witte and Rogge 2011). The reason for this concern is twofold. On the one hand, there are the numerous findings

[28] This section is based on De Witte and Rogge (2011).

in the academic literature which suggest that one or more background conditions (e.g., class size, subject matter, teacher gender, teacher experience, course grades, timing of the course) may have a significant influence on SET-scores (see, for instance, Feldman 1977; Marsh 1987, 2007; Marsh and Roche 1997; Centra and Gaubatz 2000; Marsh and Roche 2000). On the other hand, there is the practical experience from teachers themselves which indicates that some teaching environments are more conducive to high-quality teaching (and, hence, high SET-scores) while other environments make such a level of teaching more difficult.

De Witte and Rogge (2011) propose a DEA based Benefit of Doubt (BoD) approach, similar to that outlined above, for assessing academics on research, to assess them on teaching effectiveness. They construct SET-scores using a large array of single-dimensional performance indicators i (with $\mathrm{i} = 1, \ldots, \mathrm{N}$) where the weight placed on each indicator is derived through the BoD model. The conceptual starting point of BoD estimators, as we saw above, is that information on the appropriate weights can be retrieved from the observed data (i.e., letting the data speak).

The data consists of student assessments of the teacher on courses j ($\mathrm{j} = 1, \ldots, \mathrm{k}, \ldots, \mathrm{N}$) so that on course k the data on questionnaire item i, is I_i^k. For each teacher the BoD model assigns weights w_i^k to I_i^k so as to maximize the teacher's SET-score $\mathrm{SET_k}$. The model is as in (12.11) in Sect. 12.3.3.

To avoid problematic weight scenarios (zero or unrealistic weights), and to ensure the weights have intuitive appeal for teachers and students, additional weight restrictions are introduced in the basic model in the form of (12.17).

$$w_i^k \in \mathrm{W_e} \qquad \mathrm{i} = 1, \ldots, \mathrm{q} \text{ and } \mathrm{e} \in \mathrm{E} \tag{12.19}$$

where W denotes the set of permissible weight values based upon the opinion of selected stakeholders $e \in E$. For more details on this point see De Witte and Rogge (2011).

The authors further deploy the order-m method pioneered by Cazals et al. (2002) so as to estimate an *outlier-robust* SET score for each teacher. Moreover, they adapt the order-m scores so that they incorporate the exogenous environment (represented by R background characteristics $z_1, \ldots z_R$). This is done by drawing with replacement with a particular probability m observations from those observations for which $\mathrm{Z_{k,r}} \simeq \mathrm{Z}$. In particular, they create a reference group $\Upsilon^{m,z}$ from those observations which have the highest probability of being similar to the evaluated observation (similar in terms of the teaching environment in which the evaluated course was taught). The latter condition corresponds to conditioning on the exogenous characteristics $\mathrm{Z_{k,r}}$ (i.e., the teacher-related, student-related and course-related background characteristics). To do so, they smooth the exogenous characteristic Z by estimating a kernel function around $\mathrm{Z_{k,r}}$. Then they use the BoD model with the adapted reference set $\Upsilon^{m,z}$ to obtain estimates, labeled as $(\mathrm{SET_{k^m}}\left(I_i^k \middle| \mathrm{z}\right))$. These scores are not only robust to outlying observations (e.g., arising from measurement errors) but they also allow for heterogeneity arising from teacher, student and course characteristics.

De Witte and Rogge (2011) used the foregoing method to assess 69 different teachers of 112 college courses k ($k = 1,\ldots,112$) of the Faculty of Business Administration of HUB. Teachers who lecture several courses had several SET-scores, i.e. one for each evaluated course. Some 5513 students provided the feedback on the courses. The questionnaire comprised 16 statements to evaluate the multiple aspects of teacher performance. Students were asked to rate the lecturers on all items on a five-point Likert scale that corresponds to a coding rule ranging from 1 (I completely disagree) to 5 (I completely agree). The questions, covered 'Learning & Value', 'Examinations & Assignments', 'Lecture Organization', and 'Individual Lecturer Report'. For each course k ($k = 1,\ldots,112$) they calculated an average student rating I_i^k for each questionnaire item i ($i = 1,\ldots, 16$):

$$I_i^k = \frac{1}{S}\sum\nolimits_{S\in\ course\ k} I_{k,i,s} \tag{12.20}$$

where $I_{k,i,s}$ denotes the rating on question i of student s for the teacher who is lecturing course k. S is the number of students rating the course concerned. In terms of contextual variables $Z_{k,r}$, noted above, age of the teacher, gender, years of experience, whether or not he/she is a guest lecturer, whether or not the teacher received pedagogical training in the past and whether or not he/she has a doctoral degree were taken into account. Further, they included three background characteristics related to the students: the actual mean grade of the students in the class, the inequality of the distribution of the student grades (as measured by the Gini coefficient which can vary between 0 and 1, with a Gini coefficient of 0 indicating a perfectly uniform distribution and a Gini of 1 designating the exact opposite), and the response rate to the questionnaire. The latter captures the ratio of the number of people who completed the teacher evaluation questionnaire (i.e., S) to the (official) class size. Finally, two characteristics related to the course are included in the analysis: the class size and a dummy indicating whether the course is taught in the evening. They assessed teachers from three perspectives allowing progressively for the exogenous variables of teacher and student characteristics. The detailed models and findings can be found in De Witte and Rogge (2011).

12.5 Conclusion

This chapter has provided insights into the richness of DEA in education literature. By applying technical and allocative efficiency and productivity change techniques to educational data, policy relevant insights are obtained at both student level, school and system level. While this Chapter provided an overview of recent work, it is definitely not complete. De Witte and López-Torres (2015) and Johnes (2015) provide two complementary literature reviews on the efficiency in education literature. Their reviews show that many authors in various countries working with heterogeneous data sources are contributing to the literature. Despite these common efforts, there are still many aspects of efficiency in education to be explored.

De Witte and López-Torres (2015) argue that it is remarkable that the DEA (or Operations Research) literature studying education is still a distinct literature from the standard parametric 'economics of education literature'. The latter literature pays significant attention to the issue of causality, while this is not an issue in the DEA literature yet. Only few DEA studies acknowledge that the presence of endogeneity (e.g., due to omitted variable bias, measurement errors or selection bias) results in internal validity problems (notable exceptions are Ruggiero 2004; Haelermans and De Witte 2012; Cordero-Ferrera et al. 2013; Santín and Sicilia 2014). If the DEA literature on education aims to have more impact on the policy debate and on policy making, it should focus more on endogeneity and causal interpretations. The results from the DEA literature can now be easily criticised because of the lack of causal evidence. In relation to this, De Witte and López-Torres (2015) argue that the DEA literature should be more outward looking. Important developments in the economics of education literature, like experiments and quasi-experiments, have been largely ignored. There are few DEA studies that exploit experimental or quasi-experimental evidence (an interesting exception is Santín and Sicilia 2014). Yet, applying DEA to data from experiments or natural experiments in education might yield promising results. One may think of examining the efficiency of educational innovations, or changes at system level. Applying DEA to this type of data would help to bridge the gap between the DEA efficiency in education literature and the parametric efficiency in education literature.

References

Abramo G, D'Angelo CA (2014) How do you define and measure research productivity? Scientometrics 102(2):1129–1144

Abramo G, D'Angelo CA, Di Costa F (2010a) Citations versus journal impact factor as proxy of quality: could the latter ever be preferable? Scientometrics 84(3):821–833

Abramo G, D'Angelo CA, Solazzi M (2010b) National research assessment exercises: a measure of the distortion of performance rankings when labor input is treated as uniform. Scientometrics 84(3):605–619

Abramo G, Cicero T, D'Angelo CA (2012a) How important is the choice of scaling factor in standardizing citations? J Informetr 6(4):645–654

Abramo G, Cicero T, D'Angelo CA (2012b) A sensitivity analysis of researchers' productivity rankings to the time of citation observation. J Informetr 6(2):192–201

Abramo G, D'Angelo CA, Cicero T (2012c) What is the appropriate length of the publication period over which to assess research performance? Scientometrics 93(3):1005–1017

Abramo G, Cicero T, D'Angelo CA (2013a) Individual research performance: a proposal for comparing apples and oranges. J Informetr 7(2):528–539

Abramo G, D'Angelo CA, Rosati F (2013b) The importance of accounting for the number of co-authors and their order when assessing research performance at the individual level in the life sciences. J Informetr 7(1):198–208

Abramo G, D'Angelo CA, Di Costa F (2014) Variability of research performance across disciplines within universities in non-competitive higher education systems. Scientometrics 98 (2):777–795

Agasisti T, Johnes G (2009) Beyond frontiers: comparing the efficiency of higher education decision-making units across more than one country. Educ Econ 17(1):59–79

Agasisti T, Ieva F, Paganoni AM (2014) Heterogeneity, school effects and achievement gaps across Italian regions: further evidence from statistical modeling. MOX report number 07/2014. http://mox.polimi.it/it/progetti/pubblicazioni/quaderni/07-2014.pdf. Dipartimento di Matematica "F. Brioschi", Politecnico di Milano, Via Bonardi 9 – 20133 Milano

Aigner D, Lovell CAK, Schmidt P (1977) Formulation and estimation of stochastic frontier production models. J Econometrics 6:21–37

Andersen P, Petersen NC (1993) A procedure for ranking efficient units in data envelopment analysis. Manag Sci 39(10):1261–1264

Arnold VL, Bardhan IR, Cooper WW, Kumbhakar SC (1996) New uses of DEA and statistical regressions for efficiency evaluation – with an illustrative application to public secondary schools in Texas. Ann Oper Res 66(4):255–277

Athanassopoulos A, Shale EA (1997) Assessing the comparative efficiency of higher education institutions in the UK by means of data envelopment analysis. Educ Econ 5(2):117–135

Ballou D, Sanders WL, Wright P (2004) Controlling for student background in value-added assessment of teachers. J Educ Behav Stat 29(1):37–65

Banker RD, Morey RC (1986) The use of categorical variables in data envelopment analysis. Manag Sci 32(12):1613–1627

Banker RD, Janakiraman S, Natarajan R (2004) Analysis of trends in technical and allocative efficiency: an application to Texas public school districts. Eur J Oper Res 154(2):477–491

Baumol WJ, Panzar JC, Willig RD (1982) Contestable markets and the theory of industry structure. Harcourt Brace Jovanovich, London

Bessent AM, Bessent EW (1980) Determining the comparative efficiency of schools through data envelopment analysis. Educ Adm Q 16(2):57–75

Bogetoft P, Nielsen K (2005) Internet based benchmarking. Group Decis Negot 14(3):195–215

Bradley S, Johnes G, Millington J (2001) The effect of competition on the efficiency of secondary schools in England. Eur J Oper Res 135(3):545–568

Burney NA, Johnes J, Al-Enezi M, Al-Musallam M (2013) The efficiency of public schools: the case of Kuwait. Educ Econ 21(4):360–379

Camanho AS, Dyson RG (2006) Data envelopment analysis and Malmquist indices for measuring group performance. J Prod Anal 26:35–49

Caporaletti LE, Dulá JH, Womer NK (1999) Performance evaluation based on multiple attributes with nonparametric frontiers. Omega 27(6):637–645

Casu B, Thanassoulis E (2006) Evaluating cost efficiency in central administrative services in UK universities. Omega 34(5):417–426

Casu B, Shaw D, Thanassoulis E (2005) Using a group support system to aid input-output identification in DEA. J Oper Res Soc 56(12):1363–1372

Cazals C, Florens J-P, Simar L (2002) Nonparametric frontier estimation: a robust approach. J Econometrics 106(1):1–25

Centra JA, Gaubatz NB (2000) Is there gender bias in student evaluations of teaching. J High Educ 71(1):17–33

Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision making units. Eur J Oper Res 2(4):429–444

Charnes A, Cooper WW, Rhodes E (1981) Evaluating program and managerial efficiency: an application of data envelopment analysis to program follow through. Manag Sci 27 (6):668–697

Cherchye L, Kuosmanen T (2006) Benchmarking sustainable development: a synthetic meta-index approach. In: McGillivray M, Clarke M (eds) Perspectives on human development. United Nations University Press, Tokyo

Cherchye L, Moesen W, Rogge N, Van Puyenbroeck T (2007) An introduction to 'benefit of the doubt' composite indicators. Soc Indic Res 82(1):111–145

Cherchye L, De Witte K, Ooghe E, Nicaise I (2010) Efficiency and equity in private and public education: a nonparametric comparison. Eur J Oper Res 202(2):563–573

Chetty R, Friedman JN, Rockoff JE (2014) Measuring the impacts of teachers I: evaluating bias in teacher value-added estimates. Am Econ Rev 104(9):2593–2632

Coleman Report (1966) The concept of equality of educational opportunity. Johns Hopkins University/US Department of Health, Education & Welfare, Office of Education, Baltimore/Washington, DC

Cordero-Ferrera JM, Crespo-Cebada E, Pedraja-Chaparro F, Santín-González D (2011) Exploring educational efficiency divergences across Spanish regions in PISA 2006. Rev Econ Apl 19 (3):117–145

Cordero-Ferrera JM, Santín D, Sicilia G (2013) Dealing with the endogeneity problem in data envelopment analysis. MPRA 4:74–75

Costa JM, Horta IM, Guimarães N, Nóvoa MH, Cunha JFe, Sousa R (2007) icBench: a benchmarking tool for Portuguese construction industry companies. Int J Hous Sci Appl 31 (1):33–41

Crespo-Cebada E, Pedraja-Chaparro F, Santín D (2014) Does school ownership matter? An unbiased efficiency comparison for regions of Spain. J Prod Anal 41(1):153–172

Daraio C, Simar L (2005) Introducing environmental variables in nonparametric frontier models: a probabilistic approach. J Prod Anal 24(1):93–121

Daraio C, Simar L (2007a) Advanced robust and nonparametric methods in efficiency analysis: methodology and applications. Springer, Dordrecht

Daraio C, Simar L (2007b) Conditional nonparametric frontier models for convex and nonconvex technologies: a unifying approach. J Prod Anal 28(1):13–32

De Witte K, Hudrlikova L (2013) What about excellence in teaching? A benevolent ranking of universities. Scientometrics 96(1):337–364

De Witte K, Kortelainen M (2013) What explains the performance of students in a heterogeneous environment? Conditional efficiency estimation with continuous and discrete environmental variables. Appl Econ 45(17):2401–2412

De Witte, K. and López-Torres, L. (2015). Efficiency in Education. A review of literature and a way forward. Documents de treball d'economia de l'empresa – Working paper series Universitat Autonoma de Barcelona. 15/01, pp. 40. Journal of Operational Research Society. In press

De Witte K, Marques RC (2009) Capturing the environment, a metafrontier approach to the drinking water sector. Int Trans Oper Res 16(2):257–271

De Witte K, Marques RC (2010) Influential observations in frontier models, a robust non-oriented approach to the water sector. Ann Oper Res 181(1):377–392

De Witte K, Rogge N (2010) To publish or not to publish? On the aggregation and drivers of research performance. Scientometrics 85(3):657–680

De Witte K, Rogge N (2011) Accounting for exogenous influences in performance evaluations of teachers. Econ Educ Rev 30(4):641–653

De Witte K, Van Klaveren C (2014) How are teachers teaching? A nonparametric approach. Educ Econ 22(1):3–23

De Witte K, Thanassoulis E, Simpson G, Battisti G, Charlesworth-May A (2010) Assessing pupil and school performance by non-parametric and parametric techniques. J Oper Res Soc 61 (8):1224–1237

De Witte K, Rogge N, Cherchye L, Van Puyenbroeck T (2013a) Accounting for economies of scope in performance evaluations of university professors. J Oper Res Soc 64(11):1595–1606

De Witte K, Rogge N, Cherchye L, Van Puyenbroeck T (2013b) Economies of scope in research and teaching: a non-parametric investigation. Omega 41(2):305–314

Deutsch J, Dumas A, Silber J (2013) Estimating an educational production function for five countries of Latin America on the basis of the PISA data. Econ Educ Rev 36:245–262

Emrouznejad A, De Witte K (2010) COOPER-framework: a unified process for non-parametric projects. Eur J Oper Res 207(3):1573–1586
Färe R (1991) Measuring Farrell efficiency for a firm with intermediate inputs. Acad Econ Pap 19 (2):329–340
Färe R, Grosskopf S (1996a) Intertemporal production frontiers: with dynamic DEA. Kluwer Academic Publishers, Boston
Färe R, Grosskopf S (1996b) Productivity and intermediate products: a frontier approach. Econ Lett 50(1):65–70
Färe R, Karagiannis G (2013) The denominator rule for share-weighting aggregation. Mimeo
Färe R, Karagiannis G (2014) A postscript on aggregate Farrell efficiencies. Eur J Oper Res 233 (3):784–786
Färe R, Zelenyuk V (2003) On aggregate Farrell efficiencies. Eur J Oper Res 146(3):615–620
Färe R, Grosskopf S, Weber WL (1989) Measuring school district performance. Public Finance Q 17(4):409–428
Farrell M (1957) The measurement of productive efficiency. J R Stat Soc Ser A 120(3):253–281
Feldman KA (1977) Consistency and variability among college students in rating their teachers and courses: a review and analysis. Res High Educ 6(3):223–274
Flegg T, Allen D, Field K, Thurlow TW (2004) Measuring the efficiency of British universities: a multi-period data envelopment analysis. Educ Econ 12(3):231–249
Førsund FR (1993) Productivity growth in Norwegian ferries. In: Fried HO, Schmidt SS (eds) The measurement of productive efficiency: techniques and applications. Oxford University Press, New York
Fukuyama H, Weber WL (2002) Evaluating public school district performance via DEA gain functions. J Oper Res Soc 53(9):992–1003
Golany B, Storbeck JE (1999) A data envelopment analysis of the operational efficiency of bank branches. Interfaces 29(3):14–26
Goldstein H (1987) Multilevel models in educational and social research. Charles Griffin, London
Goldstein H, Huiqi P, Rath T, Hill N (2000) The use of value added information in judging school performance. Institute of Education, London
Gray J, Jesson D, Jones B (1986) The search for a fairer way of comparing schools' examination results. Res Pap Educ 1(2):91–122
Greene W (2005) Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. J Econometrics 126:269–303
Grosskopf S, Hayes KJ, Taylor LL, Weber WL (1997) Budget-constrained frontier measures of fiscal quality and efficiency in schooling. Rev Econ Stat 79:116–124
Grosskopf S, Hayes KJ, Taylor LL, Weber WL (1999) Anticipating the consequences of school reform: a new use of DEA. Manag Sci 45(4):608–620
Haegeland T (2006) School performance indicators in Norway: a background report for the OECD project on the development of value-added models in education systems. OECD, Paris
Haelermans C, De Witte K (2012) The role of innovations in secondary school performance – evidence from a conditional efficiency model. Eur J Oper Res 223(2):541–549
Hagen NT (2014) Counting and comparing publication output with and without equalizing and inflationary bias. J Informetr 8(2):310–317
Hanushek EA (1979) Conceptual and empirical issues in the estimation of educational production functions. J Hum Resour 14(3):351–388
Hanushek EA (1986) The economics of schooling: production and efficiency in public schools. J Econ Lit 24(3):1141–1177
Hanushek EA, Link S, Woessmann L (2013) Does school autonomy make sense everywhere? Panel estimates from PISA. J Dev Econ 104:212–232
Ibensoft Aps (2013) Interactive benchmarking: state-of-the-art in performance evaluation. From http://www.ibensoft.com/
Johnes G (1998) The costs of multi-product organisations and the heuristic evaluation of industrial structure. Socioecon Plann Sci 32(3):199–209

Johnes G (1999) The management of universities: Scottish Economic Society/Royal Bank of Scotland annual lecture. Scott J Polit Econ 46:502–522

Johnes J (2006a) Data envelopment analysis and its application to the measurement of efficiency in higher education. Econ Educ Rev 25(3):273–288

Johnes J (2006b) Measuring efficiency: a comparison of multilevel modelling and data envelopment analysis in the context of higher education. Bull Econ Res 58(2):75–104

Johnes J (2006c) Measuring teaching efficiency in higher education: an application of data envelopment analysis to economics graduates from UK universities 1993. Eur J Oper Res 174:443–456

Johnes J (2008) Efficiency and productivity change in the English higher education sector from 1996/97 to 2004/05. Manch Sch 76(6):653–674

Johnes G (2013) Efficiency in higher education institutions revisited: a network approach. Econ Bull 33(4):2698–2706

Johnes J (2014) Efficiency and mergers in English higher education 1996/97 to 2008/9: parametric and non-parametric estimation of the multi-input multi-output distance function. Manch Sch 82 (4):465–487

Johnes J (2015) Operational research in education. Eur J Oper Res 243(3):683–696

Johnes G, Johnes J (2009) Higher education institutions' costs and efficiency: taking the decomposition a further step. Econ Educ Rev 28(1):107–113

Johnes J, Johnes G (2013) Efficiency in the higher education sector: a technical exploration. Department for Business Innovation and Skills, London

Johnes G, Johnes J, Thanassoulis E, Lenton P, Emrouznejad A (2005) An exploratory analysis of the cost structure of higher education in England. Department for Education and Skills, London

Johnes G, Johnes J, Thanassoulis E (2008) An analysis of costs in institutions of higher education in England. Stud High Educ 33(5):527–549

Johnes J, Izzeldin M, Pappas V (2014) A comparison of performance of Islamic and conventional banks 2004-2009. J Econ Behav Organ 103:S93–S107

Johnson AL, McGinnis L (2011) Performance measurement in the warehousing industry. IIE Trans 43(3):220–230

Kao C, Hung H-T (2003) Ranking university libraries with *a posteriori* weights. Libri 53:282–289

Kao C, Hung H-T (2005) Data envelopment analysis with common weights: the compromise solution approach. J Oper Res Soc 56(10):1196–1203

Kao C, Hung H-T (2007) Management performance: an empirical study of the manufacturing companies in Taiwan. Omega 35(2):152–160

Kao C, Wu W-Y, Hsieh W-J, Wang T-Y, Lin C, Chen L-H (2008) Measuring the national competitiveness of Southeast Asian countries. Eur J Oper Res 187(2):613–628

Karagiannis G (2016) On Aggregate Composite Indicators, J Oper Res Soc (forthcoming)

Karagiannis G (2015) On structural and average technical efficiency. J Prod Anal 43(3):259–267

Karagiannis G, Lovell CAK (2016) Productivity measurement in radial Dea models with multiple constant inputs. Eur J Oper Res (fothcoming)

Karagiannis G, Paleologou SM (2014) Towards a composite public sector performance indicator. Asia-Pacific productivity conference, Brisbane, 8–11 Jul 2014

Karagiannis G, Paschalidou G (2014) Assessing effectiveness of research activity at the faculty and department level: a comparison of alternative models. Efficiency in education workshop, The Work Foundation, 19–20 Sept 2014

Kirjavainen T, Loikkanen HA (1998) Efficiency differences of Finnish senior secondary schools: an application of DEA and Tobit analysis. Econ Educ Rev 17(4):377–394

Kuosmanen T (2002) Modeling blank entries in data envelopment analysis. EconWPA working paper at WUSTL No. 0210001

Lazarsfeld PF, Henry NW (1968) Latent structure analysis. Houghton Mifflin, New York

Lin C-S, Huang M-H, Chen D-Z (2013) The influences of counting methods on university rankings based on paper count and citation count. J Informetr 7(3):611–621

Liu WB, Zhang DQ, Meng W, Li XX, Xu F (2011) A study of DEA models without explicit inputs. Omega 39(5):472–480
Lovell CAK, Pastor JT (1999) Radial DEA models without inputs or without outputs. Eur J Oper Res 118(1):46–51
Malmquist S (1953) Index numbers and indifference surfaces. Trab Estat 4:209–242
Mancebón M-J, Calero J, Choi Á, Ximénez-de-Embún DP (2012) The efficiency of public and publicly subsidized high schools in Spain: evidence from PISA-2006. J Oper Res Soc 63:1516–1533
Marsh HW (1987) Students' evaluations of university teaching: research findings, methodological issues, and directions for further research. Int J Educ Res 11:253–288
Marsh HW (2007) Students' evaluations of university teaching: dimensionality, reliability, validity, potential biases and usefulness. In: Perry RP, SMart JC (eds) The scholarship of teaching and learning in higher education: an evidence-based perspective. Springer, Dordrecht, pp 319–383
Marsh HW, Roche L (1997) Making students' evaluations of teaching effectiveness effective: the critical issues of validity, bias, and utility. Am Psychol 52(11):1187–1197
Marsh HW, Roche L (2000) Effects of grading leniency and low workload on students' evaluations of teaching, popular myth, bias, validity, or innocent bystanders? J Educ Psychol 92 (1):202–228
Mayston DJ (2003) Measuring and managing educational performance. J Oper Res Soc 54 (7):679–691
Meyer RH (1997) Value-added indicators of school performance: a primer. Econ Educ Rev 16 (3):283–301
Mizala A, Romaguera P, Farren D (2002) The technical efficiency of schools in Chile. Appl Econ 34(12):1533–1552
Muñiz M (2002) Separating managerial inefficiency and external conditions in data envelopment analysis. Eur J Oper Res 143:625–643
Ng WL (2007) A simple classifier for multiple criteria ABC analysis. Eur J Oper Res 177 (1):344–353
Ng WL (2008) An efficient and simple model for multiple criteria supplier selection problem. Eur J Oper Res 186(3):1059–1067
O'Donnell C (2012) An aggregate quantity framework for measuring and decomposing productivity and profitability change. J Prod Anal 38(3):255–272
OECD (2008) Measuring improvements in learning outcomes: best practices to assess the value-added of schools. OECD Publishing, Paris
Oliveira MA, Santos C (2005) Assessing school efficiency in Portugal using FDH and bootstrapping. Appl Econ 37:957–968
Oral M, Oukil A, Malouin J-L, Kettani O (2014) The appreciative democratic voice of DEA: a case of faculty academic performance evaluation. Socioecon Plann Sci 48(1):20–28
Orea L, Kumbhakar SC (2004) Efficiency measurement using a latent class stochastic frontier model. Empir Econ 29(1):169–183
Pastor JT, Ruiz JL, Sirvent I (2002) A statistical test for nested radial DEA models. Oper Res 50 (4):728–735
Perelman S, Santín D (2011) Measuring educational efficiency at student level with parametric stochastic distance functions: an application to Spanish PISA results. Educ Econ 19(1):29–49
Portela MCAS, Camanho AS (2007) Performance assessment of Portuguese secondary schools. Working papers de Economia number 07/2007. https://ideas.repec.org/p/cap/wpaper/072007.html, Faculdade de Economia e Gestão, Universidade Católica Portuguesa (Porto)
Portela MCAS, Camanho AS (2010) Analysis of complementary methodologies for the estimation of school value added. J Oper Res Soc 61(7):1122–1132
Portela MCAS, Thanassoulis E (2001) Decomposing school and school-type efficiency. Eur J Oper Res 132(2):357–373
Portela MCAS, Camanho AS, Borges DN (2011) BESP – benchmarking of Portuguese secondary schools. Benchmarking 18(2):240–260

Portela MCAS, Camanho AS, Borges D (2012) Performance assessment of secondary schools: the snapshot of a country taken by DEA. J Oper Res Soc 63(8):1098–1115

Portela MCAS, Camanho AS, Keshvari A (2013) Assessing the evolution of school performance and value-added: trends over four years. J Prod Anal 39(1):1–14

Raudenbush S, Bryk AS (1986) Hierarchical models for studying school effects. Sociol Educ 59 (1):1–17

Ray A (2006) School value added measures in England: a paper for the OECD project on the development of value-added models in education systems. Department for Education and Skills, London

Ray A, Evans H, McCormack T (2009) The use of national value-added models for school improvement in English schools. Rev Educ 348:47–66

Ruggiero J (1998) Non-discretionary inputs in data envelopment analysis. Eur J Oper Res 111 (3):461–469

Ruggiero J (1999) Non-parametric analysis of educational costs. Eur J Oper Res 119:605–612

Ruggiero J (2004) Performance evaluation in education: modeling educational production. In: Cooper WW, Seiford LM, Zhu J (eds) Handbook on data envelopment analysis. Kluwer Academic Publishers, Boston

Sammons P, Nuttall D, Cuttance P (1993) Differential school effectiveness: results from a reanalysis of the Inner London Education Authority's junior school project data. Br Educ Res J 19(4):381–405

Sanders WL, Saxton AM, Horn SP (1997) The Tennessee value-added assessment system: a quantitative, outcomes-based approach to educational assessment. In: Millman J (ed) Grading teachers, grading schools: is student achievement a valid evaluation measure? Corwin Press, Inc. (Sage Publications), Thousand Oaks

Santín D, Sicilia G (2014) The teacher effect: an efficiency analysis from a natural experiment in Spanish primary schools. Efficiency in education workshop, The Work Foundation, 19–20 Sept 2014

Simpson G (2005) Programmatic efficiency comparisons between unequally sized groups of DMUs in DEA. J Oper Res Soc 56(12):1431–1438

Thanassoulis E (1996) A data envelopment analysis approach to clustering operating units for resource allocation purposes. Omega 24(4):463–476

Thanassoulis E (1999) Setting achievement targets for school children. Educ Econ 7(2):101–119

Thanassoulis E (2001) Introduction to the theory and application of data envelopment analysis: a foundation text with integrated software. Kluwer Academic Publishers, Boston

Thanassoulis E, Portela MCAS (2002) School outcomes: sharing the responsibility between pupil and school. Educ Econ 10(2):183

Thanassoulis E, Portela MCAS, Despić O (2008) DEA – the mathematical programming approach to efficiency analysis. In: Fried HO, Lovell CAK, Schmidt SS (eds) The measurement of productive efficiency and productivity growth. Oxford University Press, New York

Thanassoulis E, Kortelainen M, Johnes G, Johnes J (2011) Costs and efficiency of higher education institutions in England: a DEA analysis. J Oper Res Soc 62(7):1282–1297

Thieme C, Prior D, Tortosa-Ausina E (2013) A multilevel decomposition of school performance using robust nonparametric frontier techniques. Econ Educ Rev 32:104–121

Tone K, Tsutsui M (2009) Network DEA: a slacks-based measure approach. Eur J Oper Res Soc 197:243–252

Tsionas EG (2002) Stochastic frontier models with random coefficients. J Appl Econ 17:127–147

van Vught FA, Westerheijden DF (2010) Multidimensional ranking: a new transparency tool for higher education and research. Center for Higher Education Policy Studies (CHEPS), University of Twente, Enschede

Wang Y-M, Luo Y, Lan Y-X (2011) Common weights for fully ranking decision making units by regression analysis. Expert Syst Appl 38(8):9122–9128

Zhou P, Ang BW, Poh KL (2007) A mathematical programming approach to constructing composite indicators. Ecol Econ 62:291–297

Chapter 13
Performance Benchmarking of School Districts in New York State

Thomas R. Sexton, Christie Comunale, Michael Shane Higuera, and Kelly Stickle

Abstract We used data envelopment analysis to measure the relative performance of New York State school districts in the 2011–2012 academic year and provided detailed alternative improvement pathways for each district. We found that 201 of the 624 (32.2 %) school districts with one or more high schools and 28 of the 31 (90.3 %) school districts with no high school were on the performance frontier. Using a mixed orientation, we found evidence that FTE teachers could be reduced by 8.4 %, FTE teacher support personnel could be reduced by 17.2 %, and FTE building administration and professional staff personnel could be reduced by 9.4 %. In addition, we found that the percentage of students who score 3 or 4 on the English exam could increase by 4.9 % points, 5.0 % points on the mathematics exam, and 5.8 % points on the science exam and the average graduation rate could increase by 5.4 % points.

Keywords Data envelopment analysis (DEA) • School district performance • Education policy • Education administration • Education finance • Benchmarking

13.1 Introduction

In 2011, New York State's 695 school districts (New York State Education Department n.d.) spent $53.7 billion (U.S. Census Bureau 2011, Table 6) to educate almost 2.7 million elementary and secondary pupils (U.S. Census Bureau 2011, Table 19), a cost of over $19,000 per pupil (U.S. Census Bureau 2011, Table 8). Elementary and secondary education accounts for nearly one-quarter of all state and local expenditures in New York State (U.S. Government Spending n.d.). While New York State has some excellent school districts, others struggle with poor standardized test scores and low graduation rates. Many of the reasons for the differences among school districts are widely accepted. These include differences in wealth, English proficiency, and inefficient use of resources.

T.R. Sexton, Ph.D. (✉) • C. Comunale, Ph.D. • M.S. Higuera, Ed.D. • K. Stickle
College of Business, Stony Brook University, Stony Brook, NY 11794-3775, USA
e-mail: Thomas.Sexton@StonyBrook.edu; Christie.Comunale@StonyBrook.edu; shane@sbasolutions.net; Kelly.Stickle@StonyBrook.edu

J. Zhu (ed.), *Data Envelopment Analysis*, International Series in Operations Research & Management Science 238, DOI 10.1007/978-1-4899-7684-0_13

Given the high cost of public education and its critical importance for the future of New York and the nation, it is natural for taxpayers, legislators, and administration officials to hold public education institutions accountable for producing high quality outcomes. To do so, we must measure the performance of each school district in an objective, data-informed manner. Commonly used methods for performance measurement under these circumstances are often called *benchmarking models*. When applied to school districts, a benchmark model identifies leading school districts, called *benchmark school districts*, and it facilitates the comparison of other school districts to the benchmark school districts. Non-benchmark school districts can focus on specific ways to improve their performance and thereby that of the overall statewide school system.

In this chapter, we present an appropriate benchmarking methodology, apply that methodology to measure the performance of New York State school districts in the 2011–2012 academic year, and provide detailed alternative improvement pathways for each school district.

13.2 Choosing an Appropriate Benchmarking Methodology

There are several methods used to perform benchmarking analysis. They differ in the nature of the data employed and the manner in which the data are analyzed. They also differ in their fundamental philosophies.

Some approaches compare individual units to some measure of central tendency, such as a mean or a median. For example, we might measure the financial performance of each firm within an industry by comparing its net income to the average net income of all firms in the industry. A moment's reflection reveals that large firms will outperform small firms simply due to their size and without regard to their managerial performance. We might attempt to correct for this by computing each firm's net income divided by its total assets, called the firm's return on assets. This approach is called ratio analysis, and a firm's performance might be measured by comparing its return on assets to the mean (or median) return on assets of all firms in the industry. Ratio analysis, however, assumes constant returns to scale—the marginal value of each dollar of assets is the same regardless of the size of the firm—and this may be a poor assumption in certain applications.

To avoid this assumption, we might perform a regression analysis using net income as the dependent variable and total assets as the independent variable. The performance of an individual firm would be determined by its position relative to the regression model, that is, a firm would be considered to be performing well if its net income were higher than predicted by the model given its total assets. We point out, however, that regression is a (conditional) averaging technique and measures units relative to average, rather than best, performance, and therefore does not achieve the primary objective of benchmarking.

Other approaches compare individual units to a measure of best, rather than average, performance. For example, we might modify ratio analysis by comparing a firm's return on assets to the largest return on assets of all firms in the industry. This has the advantage of revealing how much the firm needs to improve its return on assets to become a financial leader in the industry. Using such a methodology, we would encourage firms to focus on the best performers, rather than on the average performers, in its industry.

The complexity of business organizations means that no one ratio can possibly measure the multiple dimensions of a firm's financial performance. Therefore, financial analysts often report a plethora of ratios, each measuring one specific aspect of the firm's performance. The result can be a bewildering array of financial ratios requiring the analyst to piece together the ratios to create a complete, and inevitably subjective, picture of the firm's financial performance.

Fortunately, there is a methodology, called data envelopment analysis (DEA) that overcomes the problems associated with ratio analysis of complex organizations. As described in the next section, DEA employs a linear programming model to identify units called decision-making units, or DMUs whose performance, measured across multiple dimensions, is not exceeded by any other units or even any other combination of units. Cook et al. (2014) argue persuasively that DEA is a powerful "balanced benchmarking" tool in helping units to achieve best practices.

13.3 Data Envelopment Analysis

DEA has proven to be a successful tool in performance benchmarking. It is particularly well suited when measuring the performance of units along multiple dimensions, as is the case with complex organizations such as school districts. DEA has been used since the 1950s in a wide variety of applications, including health care, banking, pupil transportation, and most recently, education. DEA's mathematical development may be traced to Charnes et al. (1978), who built on the work of Farrell (1957) and others. The technique is well documented in the management science literature (Charnes et al. 1978, 1979, 1981; Sexton 1986; Sexton et al. 1986; Cooper et al. 1999), and it has received increasing attention as researchers have wrestled with problems of productivity measurement in the services and nonmarket sectors of the economy. Cooper et al. (2011) covers several methodological improvements in DEA and describes a wide variety of applications in banking, engineering, health care, and services. Emrouznejad et al. (2008) provided a review of more than 4000 DEA articles. Liu et al. (2013) use a citation-based approach to survey the DEA literature and report finding 4936 DEA papers in the literature. See deazone.com for an extensive bibliography of DEA publications as well as a DEA tutorial and DEA software.

DEA empirically identifies the best performers by forming the performance frontier based on observed indicators from all units. Consequently, DEA bases the resulting performance scores and potential performance improvements entirely

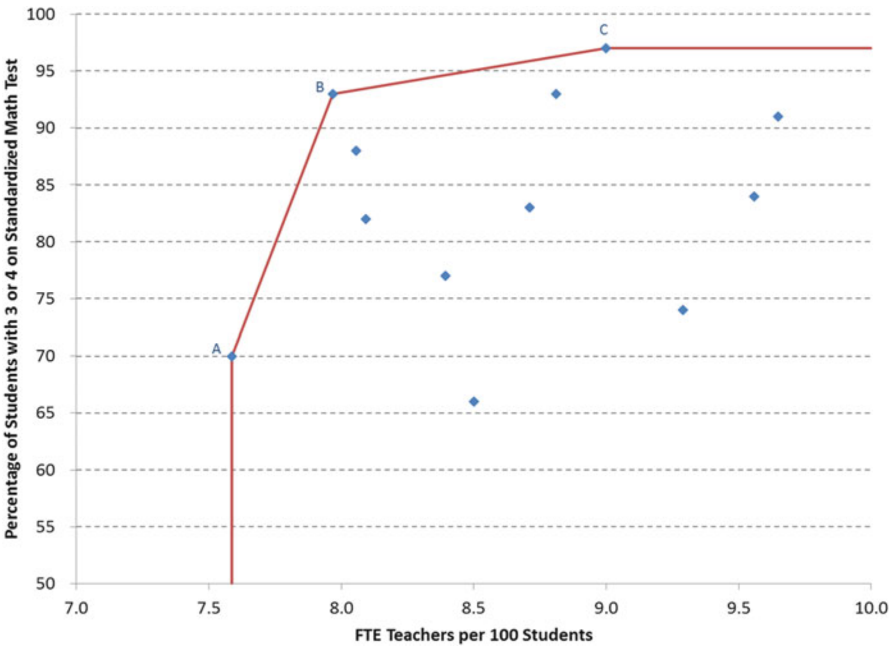


Fig. 13.1 The performance frontier for a simple example

on the actual performance of other DMUs, free of any questionable assumptions regarding the mathematical form of the underlying production function. On balance, many analysts view DEA as preferable to other forms of performance measurement.

Figures 13.1 and 13.2 illustrate the performance frontier for a simple model of school districts. We can use this simple model, which is clearly inadequate for capturing the complexity of school districts, to demonstrate the fundamental concepts of DEA. In this model, we assume that each school district employs only one type of resource, full-time equivalent (FTE) teachers, and prepares students for only one type of standardized test, mathematics at the appropriate grade level, measured as the percentage of students who score at a given level or higher. Each school district is represented by a point in the scatterplot.

In Fig. 13.1, school districts A, B, and C define the performance frontier. In each case, there is no school district or weighted average of school districts that has fewer FTE teachers per 100 students and has a higher percentage of students who scored 3 or 4 on the standardized mathematics test. Such school districts, if they existed, would lie to the Northwest of A, B, or C, and no such districts, or straight lines between any two districts, exists.

School district D, in Fig. 13.2, does not lie on the performance frontier and therefore its performance can improve. In principle, D can choose to move anywhere on the performance frontier. If school district D chooses to focus on resource reduction without test performance change, it would move to the left, reaching the

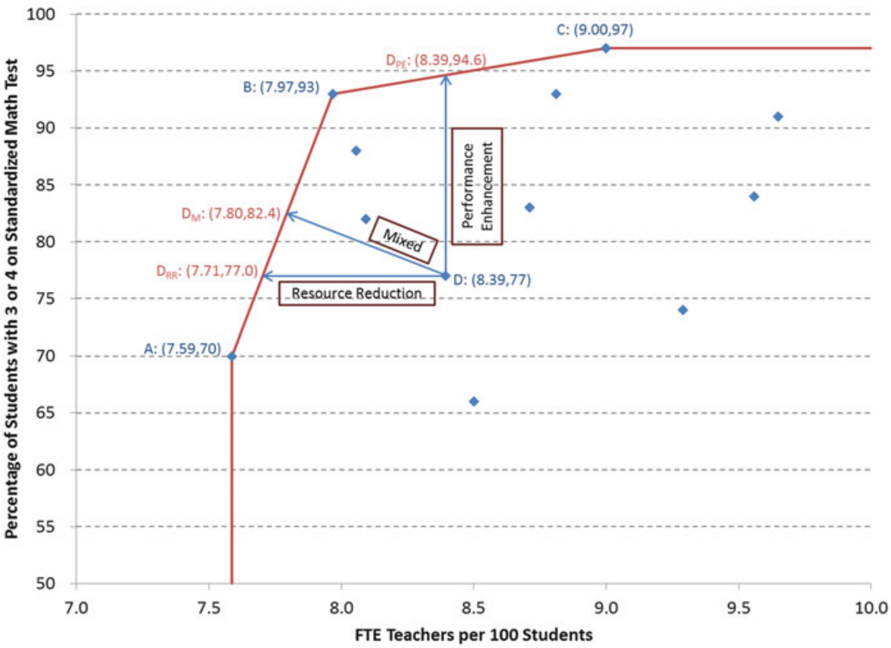


Fig. 13.2 Several ways for school district D to move to the performance frontier

performance frontier at point D_{RR}. This move would require a reduction from 8.39 to 7.71 FTE teachers per 100 students. If school district D enrolls 10,000 students, this reduction would be from 839 to 771 teachers, a percentage reduction of 8.1 %. We refer to this strategy as the *resource reduction orientation*.

If school district D chooses to focus on performance enhancement without resource reduction, it would move upward, reaching the performance frontier at point D_{PE}. This move would require 94.6 % of its students to score 3 or 4 on the standardized mathematics test, up from 77 %. If 1000 students in school district D sat for the standardized mathematics test, students scoring 3 or 4 would increase would from 770 to 946, or by 22.9 %. We refer to this strategy as the *performance enhancement orientation*.

School district D might prefer an intermediate approach that includes both resource reduction and performance enhancement and move to point D_M. This entails both a reduction in FTE teachers per 100 students from 8.39 to 7.80 and an increase in the percentage of students who score 3 or 4 on the standardized mathematics test from 77 to 82.4 %. If school district D enrolls 10,000 students, this reduction would be from 839 to 780 teachers, or by 7.0 %, and an increase in students scoring 3 or 4 from 770 to 824, or 7.0 %. We refer to this strategy as the *mixed orientation*. The mixed orientation has the feature that the percentage decrease in each resource equals the percentage increase in each performance measure.

The three points D_{RR}, D_{PE}, and D_M are called *targets* for school district D because they represent three possible goals for D to achieve to reach the performance frontier. School district D can choose its target anywhere on the performance frontier, but these three points represent reasonable reference points for D as it improves its overall performance.

Of course, this model does not consider other resources used by school districts such as teacher support personnel and other staff, nor does it consider standardized test scores in science or English. It also ignores graduation rates in school districts with one or more high schools. Moreover, it does not recognize differences in important district characteristics such as the number of elementary and secondary students, the percentage of students who qualify for free or reduced price lunch or who have limited English proficiency, or the district's combined wealth ratio.

When other measures are included in the model, we can no longer rely on a simple graphical method to identify a school district's target school district. For this purpose, we rely on the linear programming model that we describe in detail in the Appendix. Nonetheless, the target school district will have the same basic interpretation. Relative to the school district in question, the target school district consumes the same or less of each resource, its students perform the same or better on each standardized test, its graduation rate is at least as high (if applicable), it educates the same number or more students, and it operates under the same or worse district characteristics.

13.4 A DEA Model for School District Performance in New York State

To apply the DEA methodology to measure the performance of New York State school districts, we began by identifying three categories of important school district measurements. They were:

- resources consumed;
- performance measures; and
- district characteristics.

We defined the resources consumed as:

- FTE teachers;
- FTE teacher support (teacher assistants + teacher aides); and
- building administration and professional staff (principals + assistant principals + other professional staff + paraprofessionals).

For school districts with no high school, we defined the performance measures as:

- percentage of students scoring at or above level 3 on ELA grade 6;
- percentage of students scoring at or above level 3 on math grade 6; and
- percentage of students scoring at or above level 3 on science grade 4.

For school districts with one or more high schools, we defined the performance measures as:

- total cohort results in secondary-level English after 4 years of instruction: percentage scoring at levels 3–4;
- total cohort results in secondary-level math after 4 years of instruction: percentage scoring at levels 3–4;
- grade 8 science: percentage scoring at levels 3–4 all students; and
- 4-year graduation rate as of August.

We defined the district characteristics as:

- number of elementary school students;
- number of secondary school students;
- percentage of students with free or reduced price lunch;
- percentage of students with limited English proficiency; and
- school district's combined wealth ratio.

We recognize that other choices of variables are possible. We use this particular set of variables because it captures a reasonable range of resources consumed, performance dimensions to be measured, and district characteristics to be taken into account. Other variables may be added if statewide data are available for every school district. Our objective is to illustrate the model and its ability to provide school districts with useful feedback for strategic planning and other purposes.

We consider all three possible orientations. The resource reduction orientation seeks to reduce resource consumption as much as possible while maintaining performance measures at their current levels. The performance enhancement orientation seeks to improve performance measures as much as possible while maintaining resource consumption at current levels. The mixed orientation seeks to improve performance measures and reduce resource consumption simultaneously in a balanced way.

We present the results of all three orientations to provide school district administrators with alternative options for reaching the performance frontier. One district might elect to focus on resource reduction; another might opt for increases in test scores and graduation rate, while a third might prefer a blended strategy that combines these two objectives. Moreover, there are infinitely many points on the performance frontier toward which a district may move; the three that we present are designed to highlight three possible alternatives.

We point out that the performance frontier is unaffected by the choice of orientation. Any district that lies on the performance frontier in one orientation will also lie on it in any other orientation. Orientation only determines the location of the target district on the performance frontier.

13.5 Data and Results

We obtained complete data for 624 public school districts with one or more high schools and 31 public school districts with no high school for the academic year 2011–2012. Complete data were unavailable for certain districts. All data were obtained from the New York State Education Department.

13.6 Results for Three Example Districts

Table 13.1 shows the results for three districts based on the model described above. These districts were selected to illustrate the manner in which the model results can be presented to school districts and how they might be interpreted.

School district A would reduce all three resources by 18.3 % using the resource reduction orientation and by 4.0 % under the mixed orientation, but would not reduce any resources under the performance enhancement orientation. Improvements in English and science would be virtually the same using all three orientations (in the range of 4 %) but the improvements in math and graduation rate are notably higher using either the performance enhancement or mixed orientations. The message for school district A is that it can raise all three test measures by about 4 % and graduation rate by about 8% with little or no reduction in resources. Alternatively, it can improve English and science (but not math) by about 4% and graduation rate by 4–5 % even with significant resource reductions. The choice of strategy would be influenced by many other factors not reflected in the model.

School district B can reduce its FTE teachers by at least 6.9 % but its greater opportunity lies in teacher support, which it can reduce by at least 27.4 %. Despite these reductions, it can improve English by almost 7% and math by almost 4 %.

Table 13.1 Results for three example districts under three orientations (in percentages)

Dist	Orientation	FTE teachers	FTE teacher support	Bld Adm and prof staff	Secondary level English (%)	Secondary level math (%)	Grade 8 science (%)	Grad rate (%)
A	Res red	81.7	81.7	81.7	103.9	100.0	103.8	104.6
	Perf enhan	100.0	100.0	100.0	104.3	**104.3**	104.3	**108.5**
	Mixed	96.0	96.0	96.0	104.3	**104.0**	104.0	**108.2**
B	Res red	90.2	**65.8**	90.2	105.3	101.5	100.0	100.0
	Perf enhan	93.1	**72.6**	100.0	106.8	103.8	101.8	101.8
	Mixed	92.8	**72.6**	98.4	106.7	103.6	101.6	101.6
C	Res red	99.7	99.7	99.7	101.1	**113.8**	100.0	100.9
	Perf enhan	100.0	100.0	100.0	101.1	**113.8**	100.1	101.1
	Mixed	99.9	99.9	99.9	101.1	**113.8**	100.1	101.0

School district C is performing very well regardless of orientation with the exception of math, which it can improve by almost 14 %.

13.7 Statewide Results

We found no evidence that 201 of the 624 (32.2 %) school districts with one or more high schools can reduce resource consumption or improve performance. The same statement applies to 28 of the 31 (90.3 %) school districts with no high school. Put another way, each of these school districts serves as its own target school district—none of these school districts can simultaneously reduce each of its resources and improve each of its performance measures while operating under the same district characteristics.

13.8 Districts with One or More High Schools

The 624 school districts with one or more high schools employed 126,470 FTE teachers, 33,035 FTE teacher support personnel, and 25,492.5 FTE building administration and professional staff in the academic year 2011–2012. The average percentage of students who scored 3 or 4 on the English exam was 84.4 %; on the mathematics exam, the average was 86.0 %, and on the science exam, the average was 81.6 %. The average graduation rate was 84.2 %. See Table 13.2.

Using a mixed orientation, we found evidence that the number of FTE teachers can be reduced by 8.4 %, the number of FTE teacher support personnel can be reduced by 17.2 %, and the number of FTE building administration and professional staff personnel can be reduced by 9.4 %. In addition, that the average[1] percentage of students who score 3 or 4 on the English exam can rise by 4.9 % points, by 5.0 % points on the mathematics exam, and by 5.8 % points on the science exam. Moreover, the average[2] graduation rate can rise by 5.4 % points.

Using a resource reduction orientation, we found evidence that the number of FTE teachers can be reduced by 19.1 %, the number of FTE teacher support personnel can be reduced by 22.3 %, and the number of FTE building administration and professional staff personnel can be reduced by 19.3 %. In addition, the average percentage of students who score 3 or 4 on the English exam can rise by 2.2 % points, by 2.4 % points on the mathematics exam, and by 3.7 % points on the science exam. Moreover, the average graduation rate can rise by 2.3 % points.

Finally, using a performance enhancement orientation, we found evidence that the number of FTE teachers can be reduced by 5.7 %, the number of FTE teacher

[1] These are unweighted averages and therefore they do not represent the statewide percentages.

[2] See previous footnote.

Table 13.2 Data and statewide results for all three orientations for school districts with one or more high schools

	FTE teachers	FTE teacher support	Building admin and prof staff	Secondary level English (%)	Secondary level math (%)	Grade 8 science (%)	Grad rate (%)
Actual	126,470	33,035	25,493	84.4	86.0	81.6	84.2
Mixed orientation							
Target	115,812	27,359	23,091	89.3	91.0	87.4	89.6
Change	10,658	5676	2402	4.9	5.0	5.8	5.4
% Change	8.4	17.2	9.4	5.8	5.8	7.1	6.4
Resource reduction orientation							
Target	**102,314**	**25,653**	**20,567**	86.7	88.4	85.3	86.5
Change	**24,156**	**7382**	**4925**	2.2	2.4	3.7	2.3
% Change	**19.1**	**22.3**	**19.3**	2.6	2.8	4.5	2.7
Performance enhancement orientation							
Target	119,311	27,913	23,687	**89.7**	**91.3**	**87.6**	**89.9**
Change	7159	5122	1805	**5.3**	**5.3**	**6.0**	**5.7**
% Change	5.7	15.5	7.1	**6.3**	**6.1**	**7.3**	**6.8**

support personnel by 15.5 %, and the number of FTE building administration and professional staff personnel by 7.1 %. In addition, the average percentage of students who score 3 or 4 on the English exam can rise by 5.3 % points, by 5.3 % points on the mathematics exam, and by 6.0 % points on the science exam. Moreover, the average graduation rate can rise by 6.8 % points.

Figures 13.3, 13.4, and 13.5 illustrate the potential improvements in the three resource categories. For districts that lie on the diagonal of one of these graphs, there is no evidence that they could reduce their use of this resource category. Other districts have the potential to reduce resource consumption by the amount that they lay below the diagonal.

Figures 13.6, 13.7, 13.8, and 13.9 illustrate the potential improvements in the four performance measures. For districts that lie on the diagonal of one of these graphs, there is no evidence that they could improve their performance in this dimension. Other districts have the potential to improve by the amount that they lay above the diagonal.

Figure 13.10 shows the histograms of the school districts for each of the three factor performances associated with the resources, excluding those districts for which no improvement is possible. Figure 13.11 shows the histograms of the school districts for each of the four factor performances associated with the performance measures, again excluding those for which no improvement is possible.

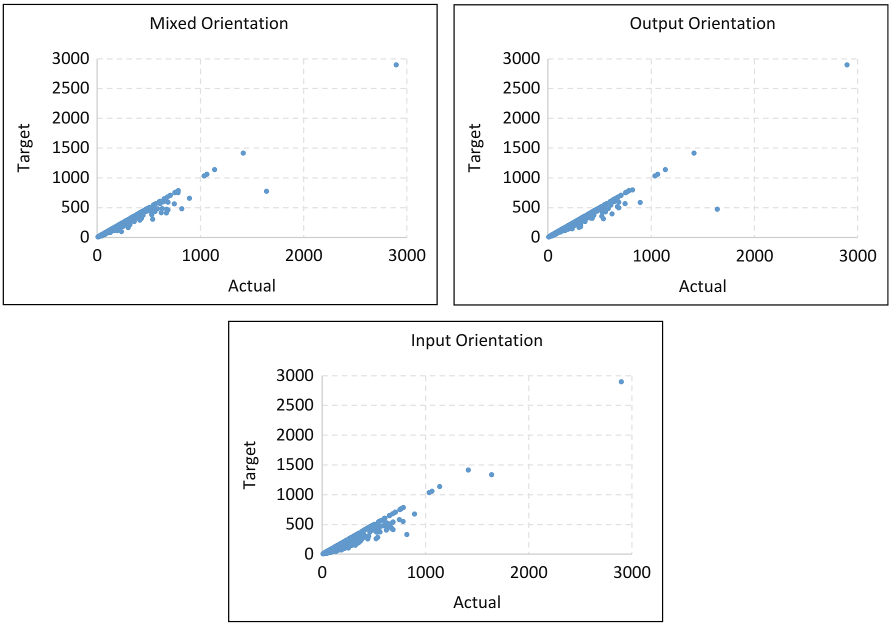


Fig. 13.3 Target vs. actual FTE teachers under each of the three orientations for school districts with at least one high school

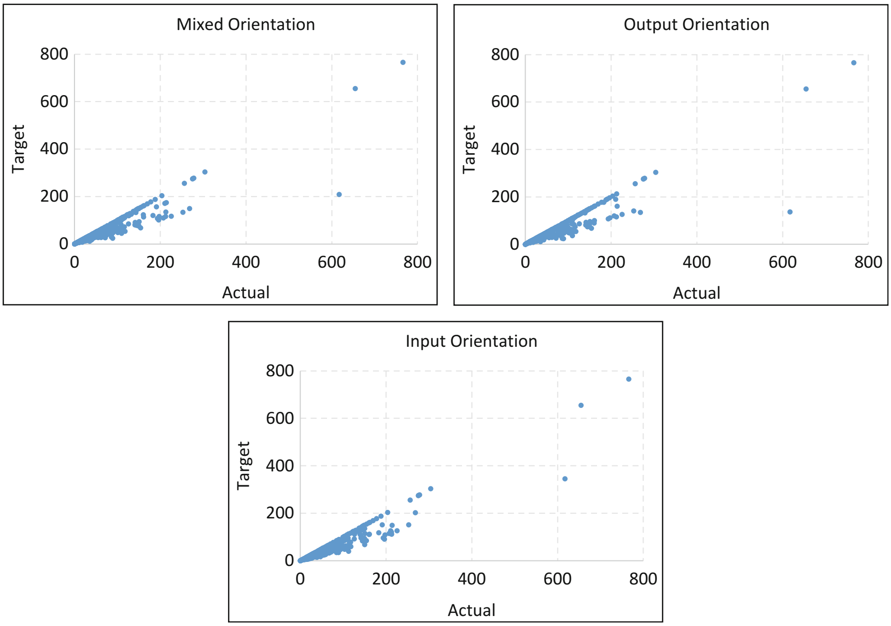


Fig. 13.4 Target vs. actual FTE teacher support under each of the three orientations for school districts with at least one high school

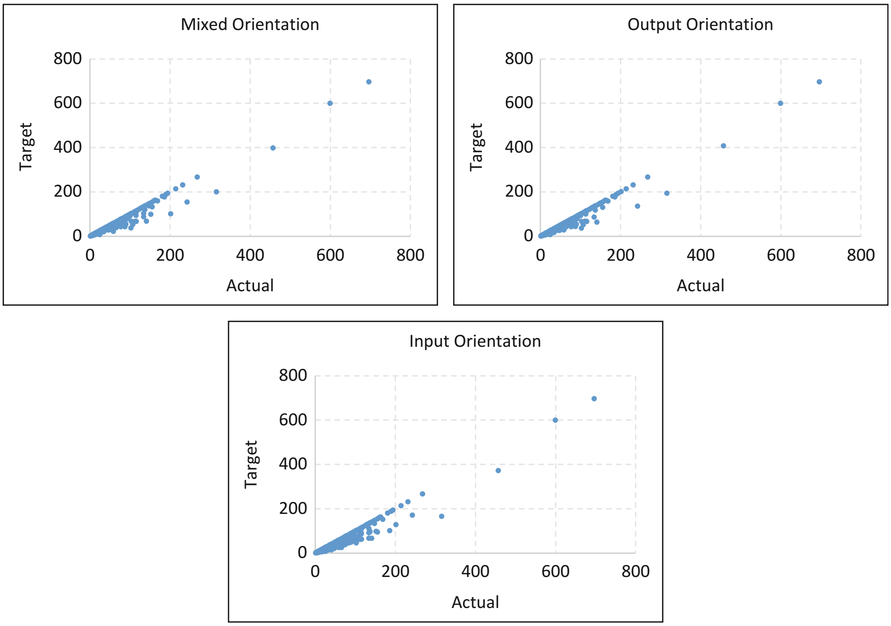


Fig. 13.5 Target vs. actual FTE building and administrative professional staff under each of the three orientations for school districts with at least one high school

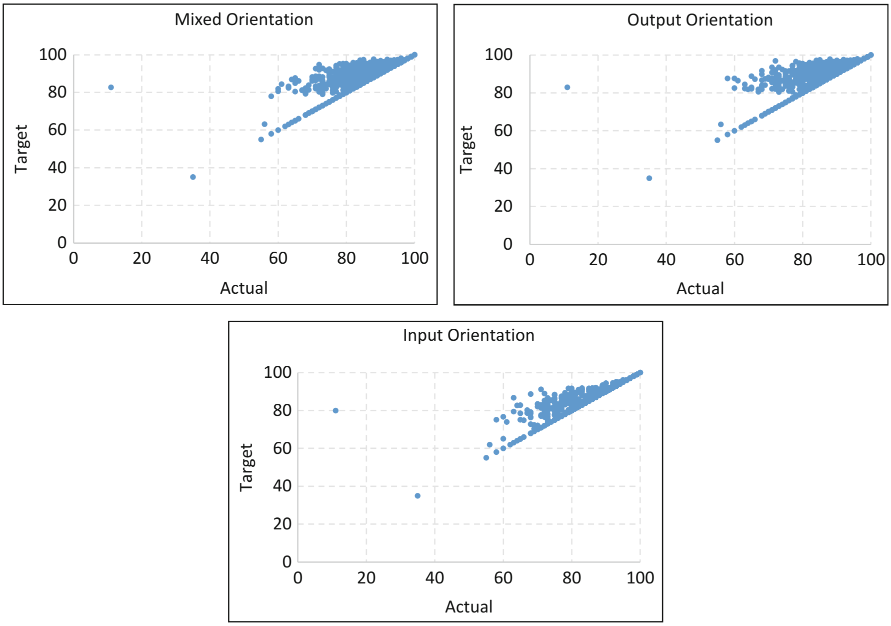


Fig. 13.6 Target vs. actual percentage of students scoring 3 or 4 on the secondary level English standardized test under each of the three orientations for school districts with at least one high school

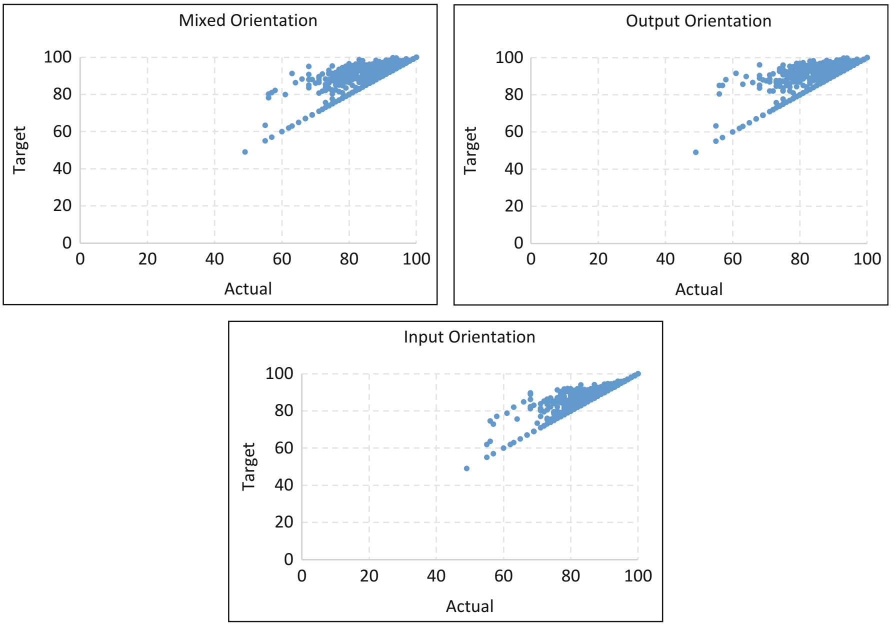


Fig. 13.7 Target vs. actual percentage of students scoring 3 or 4 on the secondary level mathematics standardized test under each of the three orientations for school districts with at least one high school

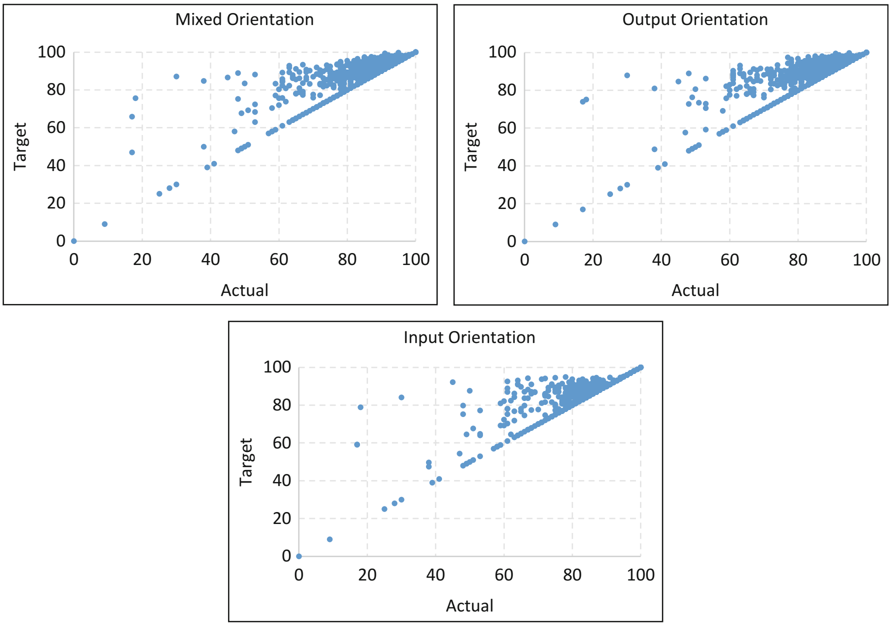


Fig. 13.8 Target vs. actual percentage of students scoring 3 or 4 on the grade 8 science standardized test under each of the three orientations for school districts with at least one high school

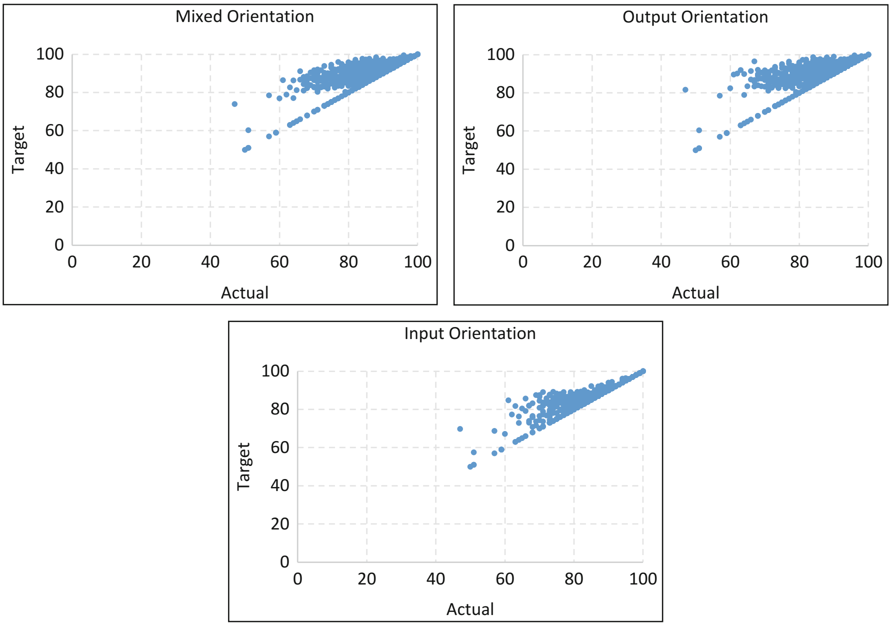


Fig. 13.9 Target vs. actual percentage of 4-year graduation rate under each of the three orientations for school districts with at least one high school

13.9 Districts Without a High School

The 31 school districts with no high school employed 2233 FTE teachers, 762 FTE teacher support personnel, and 416 FTE building administration and professional staff in the academic year 2011–2012. The average percentage of students who scored 3 or 4 on the English exam was 84.4 %; on the mathematics exam, the average was 86.0 %, and on the science exam, the average was 81.6 %. See Table 13.3.

Using a mixed orientation, we found evidence that the number of FTE teachers can be reduced by 0.2 %, the number of FTE teacher support personnel by 4.3 %, and the number of FTE building administration and professional staff personnel by 3.3 %. In addition, the average percentage of students who score 3 or 4 on the English exam can rise by 0.4 % points, by 0.9 % points on the mathematics exam, and by 0.3 % points on the science exam.

Using a resource reduction orientation, we found evidence that the number of FTE teachers can be reduced by 0.8 %, the number of FTE teacher support personnel by 4.6 %, and the number of FTE building administration and professional staff personnel by 4.8 %. In addition, the average percentage of students who score 3 or 4 on the English exam can rise by 0.6 % points, by 0.6 % points on the mathematics exam, and by 0.0 % points on the science exam.

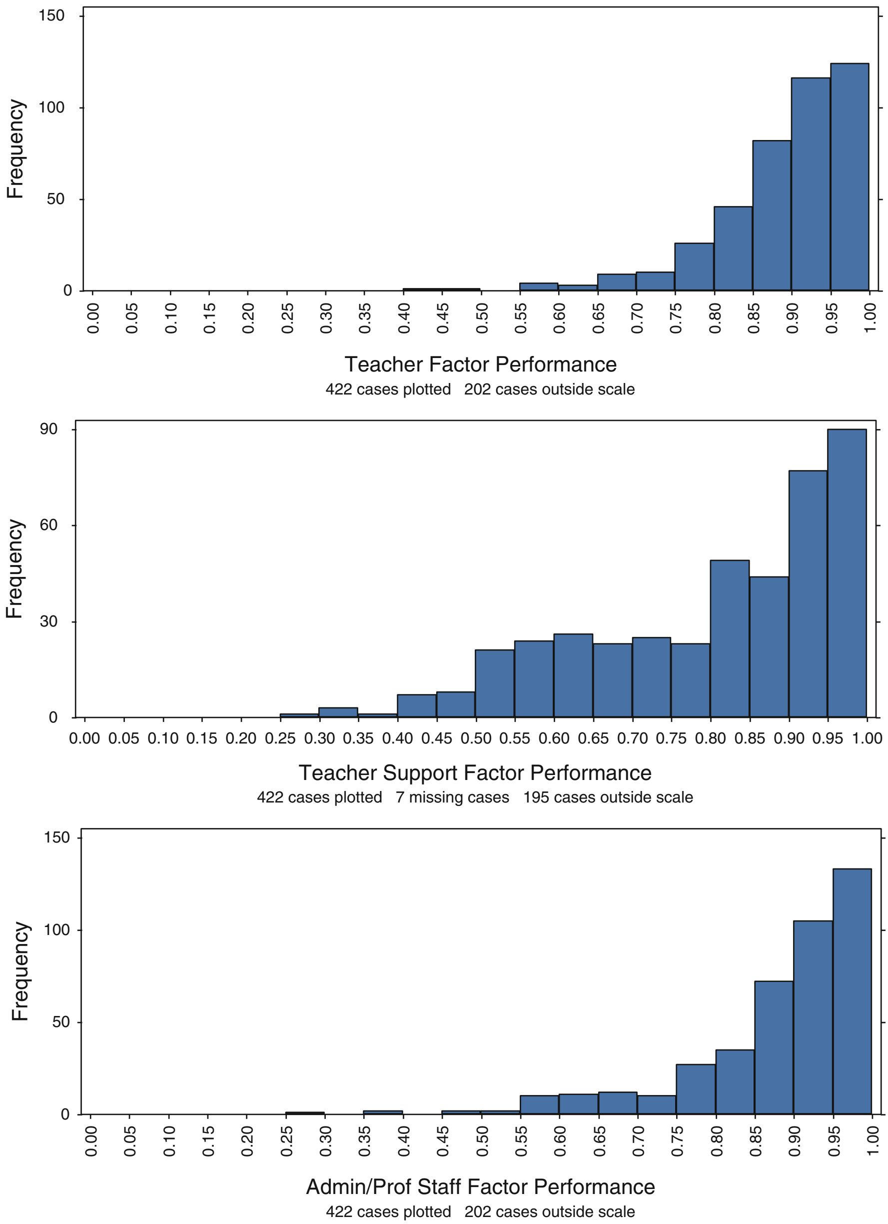


Fig. 13.10 Histograms of the school districts with at least one high school for each of the three factor performances associated with the resources, excluding those districts for which no improvement is possible

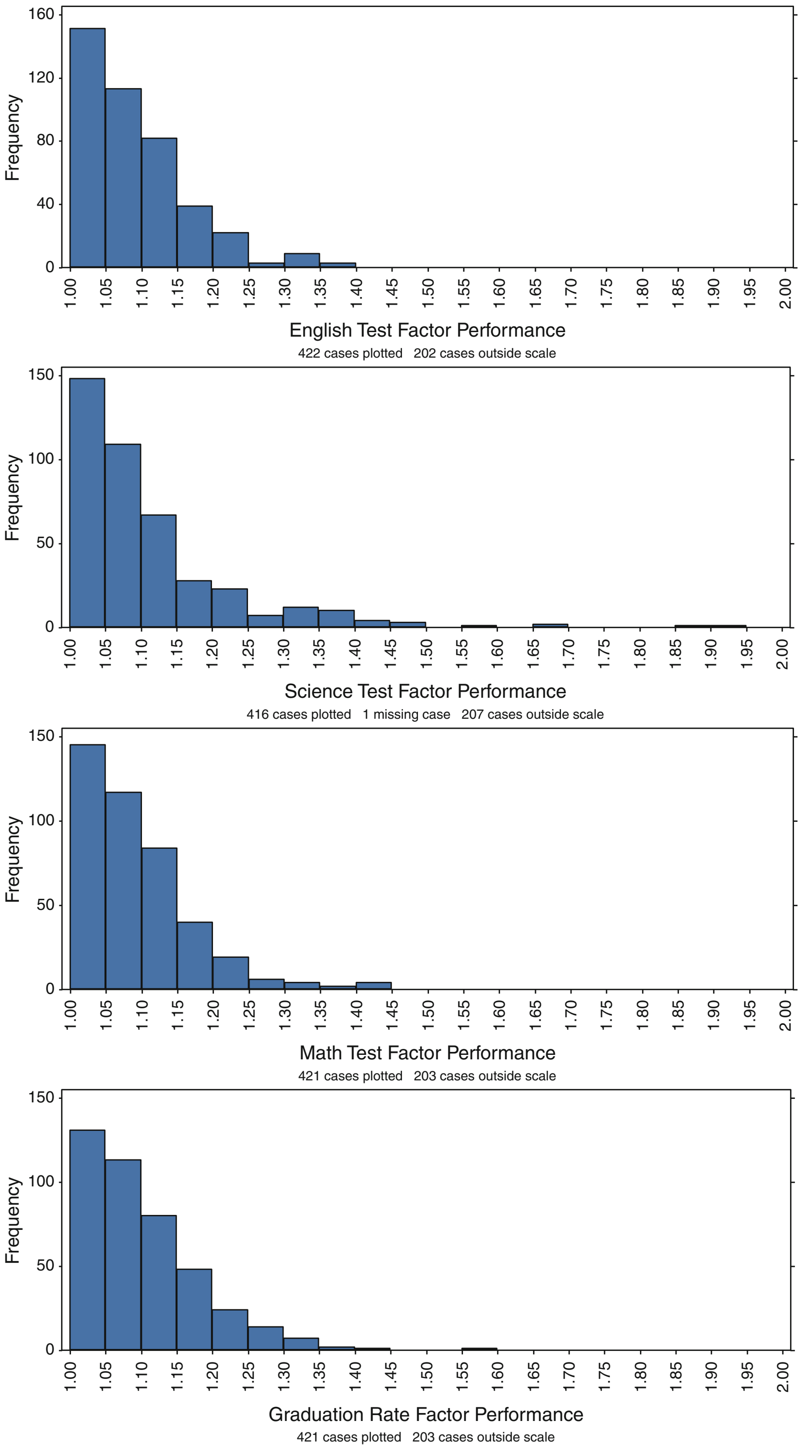


Fig. 13.11 Histograms of the school districts with at least one high school for each of the four factor performances associated with the performance measures, excluding those for which no improvement is possible

Table 13.3 Statewide results for all three orientations for school districts without a high school

	FTE teachers	FTE teacher support	Building admin and prof staff	Grade 6 ELA (%)	Grade 6 math (%)	Grade 4 science (%)
Actual	2233	762	417	77.7	83.1	94.6
Mixed orientation						
Target	2228	729	403	78.0	83.8	94.8
Change	5	33	14	0.3	0.7	0.3
% Change	0.2	4.3	3.3	0.4	0.9	0.3
Resource reduction orientation						
Target	**2216**	**727**	**397**	78.2	83.6	94.6
Change	**17**	**35**	**20**	0.5	0.5	0.0
% Change	**0.8**	**4.6**	**4.8**	0.6	0.6	0.0
Performance enhancement orientation						
Target	2233	729	404	**78.1**	**83.9**	**94.9**
Change	–	33	13	**0.3**	**0.7**	**0.3**
% Change	0.0	4.3	3.0	**0.4**	**0.9**	**0.3**

Finally, using a performance enhancement orientation, we found evidence that the number of FTE teachers can be reduced by 0.0 %, the number of FTE teacher support personnel by 4.3 %, and the number of FTE building administration and professional staff personnel by 3.0 %. In addition, the average percentage of students who score 3 or 4 on the English exam can rise by 0.4 % points, by 0.9 % points on the mathematics exam, and by 0.3 % points on the science exam.

13.10 Implementation

We reiterate that other choices of variables are possible. An important first step is for the school districts and the New York State Education Department (NYSED) to work together to modify this model as necessary. For example, the current model does not include data on Regents exam scores. In principle, the only requirement is that complete data exists for all school districts for the specified school year. In addition, it is important to provide a complete data set so that all school districts, especially those in New York City, can be included. This data set needs to be compiled for the latest school year for which complete data are available.

The NYSED would need to determine the distribution of model results. Perhaps the initial distribution during a pilot phase should be restricted to the school districts and NYSED. This would allow school districts the opportunity to understand the full meaning of their own results better and to begin to incorporate the results into their operations and planning. The pilot phase would also allow school districts and NYSED to suggest further improvements in the model.

Ultimately, the model can serve as a key element in a quality improvement cycle. By providing direct feedback to each school district about its performance along multiple dimensions, it supports school district decisions about how to improve and allows them to demonstrate that their decisions have in fact had the desirable effects.

13.11 Conclusions

We have presented a flexible model that allows school districts and NYSED to measure school district performance throughout New York State. The model provides multiple, mathematically-derived performance measures that allow school districts to detect specific areas for improvement. The model also enables NYSED to identify school districts that are the top performers in the state and others that most require improvement.

The results of a preliminary version of the model applied to data from the 2011–2012 school year shows that approximately one-third of the school districts in New York State are performing as well as can be expected given their local school district characteristics. Another 26.8–42.3 %, depending on the specific resource or performance measure, can improve by no more than 10 %.

Nonetheless, substantial statewide improvements are possible. Using the mixed orientation, for example, if every school district were to match to its target, New York State would have between 8 and 17 % fewer personnel, 6–7 % more students scoring 3 or 4 on standardized tests, and 6% more students graduating within 4 years.

Public education is critically important to the future of New York State and the nation. This model offers the potential to support public school education leaders in recognizing where improvements are possible and in taking appropriate action to implement those improvements.

Appendix: The Mathematics of the DEA Model

We use two slightly different DEA models in this chapter, one for school districts with one or more high schools, and one for school districts without a high school. The differences lie in the performance measures (different points at which test scores are measured, and no graduation rate for school districts with no high school). In addition, each model is employed with three different orientations (resource reduction, performance enhancement, and mixed). The text that follows describes the model for school districts with one or more high schools.

Let $n = 624$ be the number of school districts to be analyzed. The DEA literature refers to units under analysis as *decision-making units*, or DMUs. Let X_{ij} be amount of resource i consumed by DMU j, for $i = 1, 2, 3$, and $j = 1, 2, \ldots, 624$. In particular, let X_{1j} be the FTE teachers in DMU j, let X_{2j} be the FTE teacher support in DMU j, and let X_{3j} be the FTE building administration and professional staff in DMU j.

Let Y_{rj} be performance measure r achieved by DMU j, for $r = 1, 2, 3, 4$ and $j = 1, 2, \ldots, 624$. In particular, let Y_{1j} be the percentage of students scoring at levels 3 or 4 in secondary-level English after 4 years of instruction in DMU j, let Y_{2j} be the percentage of students scoring at levels 3 or 4 in secondary-level math after 4 years of instruction in DMU j, let Y_{3j} be the percentage of students scoring at levels 3 or 4 in Grade 8 Science in DMU j, and let Y_{4j} be the 4-year graduation rate as of August in DMU j, for $j = 1, 2, \ldots, 624$.

Let S_{kj} be the value of site characteristic k at DMU j, for $k = 1, 2, 3, 4, 5$ and $j = 1, 2, \ldots, 624$. In particular, let S_{1j} be the number of elementary school students in DMU j, let S_{2j} be the number of secondary school students in DMU j, let S_{3j} be the percentage of students with free or reduced price lunch in DMU j, let S_{4j} be the percentage of students with limited English proficiency in DMU j, and let S_{5j} be the combined wealth ratio in DMU j, for $j = 1, 2, \ldots, 624$.

The Resource Reduction DEA Model

The resource reduction DEA model with variable returns to scale, for DMU d, $d = 1, 2, \ldots, 624$, is below. We must solve $n = 624$ linear programs to perform the entire DEA.

Min E_d	(13.1)	
subject to		
$\sum_{j=1}^{n} \lambda_j X_{1j} \leq E_d X_{1d}$	(13.2a)	FTE teachers
$\sum_{j=1}^{n} \lambda_j X_{2j} \leq E_d X_{2d}$	(13.2b)	FTE teacher support
$\sum_{j=1}^{n} \lambda_j X_{3j} \leq E_d X_{3d}$	(13.2c)	Building administration and professional staff
$\sum_{j=1}^{n} \lambda_j Y_{1j} \geq Y_{1d}$	(13.3a)	Secondary level English (%)
$\sum_{j=1}^{n} \lambda_j Y_{2j} \geq Y_{2d}$	(13.3b)	Secondary level math (%)
$\sum_{j=1}^{n} \lambda_j Y_{3j} \geq Y_{3d}$	(13.3c)	Grade 8 science (%)
$\sum_{j=1}^{n} \lambda_j Y_{4j} \geq Y_{4d}$	(13.3d)	Graduation rate (%)
$\sum_{j=1}^{n} \lambda_j S_{1j} \geq S_{1d}$	(13.4a)	Number of elementary school students

(continued)

$\sum_{j=1}^{n} \lambda_j S_{2j} \geq S_{2d}$	(13.4b)	Number of secondary school students
$\sum_{j=1}^{n} \lambda_j S_{3j} \geq S_{3d}$	(13.4c)	Percentage of students with free or reduced price lunch
$\sum_{j=1}^{n} \lambda_j S_{4j} \geq S_{4d}$	(13.4d)	Percentage of students with limited English proficiency
$\sum_{j=1}^{n} \lambda_j S_{5j} \leq S_{5d}$	(13.4e)	School district's combined wealth ratio
$\sum_{j=1}^{n} \lambda_j = 1$	(13.5)	Variable returns to scale
$\lambda_j \geq 0 \, for \, j = 1,\ 2,\ \ldots,\ 624$	(13.6)	Nonnegativity
$E_d \geq 0$	(13.7)	Nonnegativity

We observe that setting $\lambda_d = 1$, $\lambda_j = 0$ for $j \neq d$, and $E_d = 1$ is a feasible, but not necessarily optimal, solution to the linear program for DMU d. This implies that $E_d^{\ *}$, the optimal value of E_d, must be less than or equal to 1. The optimal value, $E_d^{\ *}$, is the *overall efficiency* of DMU j. The left-hand-sides of (13.2)–(13.4) are weighted averages, because of (13.5), of the resources, performance measures, and site characteristics, respectively, of the 524 DMUs. At optimality, that is with the λ_j replaced by $\lambda_j^{\ *}$, we call the left-hand-sides of (13.2a)–(13.4e) the *target resources*, *target performance measures*, and *target site characteristics*, respectively, for DMU d.

Equations (13.2a)–(13.2c) imply that each target resource will be less than or equal to the actual level of that resource at DMU d. Similarly, (13.3a)–(13.3d) imply that each target performance measure will be greater than or equal to the actual level of that performance measure at DMU d.

The nature of each site characteristic inequality in (13.4a)–(13.4e) depends on the manner in which the site characteristic influences efficiency. Equations (13.4a)–(13.4d) correspond to unfavorable site characteristics (larger values imply a greater need for resources to obtain a given performance level, on average); therefore, we use the greater-than-or-equal to sign. Equation (13.4e) corresponds to a favorable site characteristic (larger values imply a lesser need for resources to obtain a given performance level, on average); therefore we use the less-than-or-equal to sign. Thus, (13.4a)–(13.4e) imply that the value of each target site characteristic will be the same as or worse than the actual value of that site characteristic at DMU d.

Thus, the optimal solution to the linear program for DMU d identifies a hypothetical target DMU d^* that, relative to DMU d, (a) consumes the same or less of every resource, (b) achieves the same or greater level of every performance measure, and (c) operates under the same or worse site characteristics. Moreover, the objective function expressed in (13.1) ensures that the target DMU d^* consumes resources levels that are reduced as much as possible in across-the-board percentage terms.

Of course, to proceed we must assume that a DMU could in fact operate exactly as does DMU d^*. In the theory of production, this is the assumption, made universally by economists, that the production possibility set is convex. In this context, the *production possibility set* is the set of all vectors $\{\boldsymbol{X}_i, \boldsymbol{Y}_r | \boldsymbol{S}_k\}$ of resources, performance measures, and site characteristics such that it is possible for a DMU to use resource levels $\boldsymbol{X}_i$ to produce performance measures $\boldsymbol{Y}_r$ under site characteristics $\boldsymbol{S}_k$. The convexity assumption assures that DMU d^* is feasible and that it is reasonable to expect that DMU d could modify its performance to match that of d^*.

We use the Premium Solver Pro© add-in (Frontline Systems, Inc., Incline Village, NV) in Microsoft Excel© to solve the linear programs. We use a macro written in Visual Basic for Applications© (VBA) to solve the 624 linear programs sequentially and save the results within the spreadsheet. Both the Basic Solver© and VBA© are available in all versions of Microsoft Excel©. However, the Basic Solver© is limited to 200 variables and 100 constraints, which limits the size of the problems to no more than 199 DMU and no more than 99 resources, performance measures, and site characteristics combined. We use the Premium Solver Pro©, available from Frontline Systems, Inc., for this application.

The Performance Enhancement DEA Model

The performance enhancement DEA model with variable returns to scale, for DMU d, $d = 1, 2, \ldots, 624$, is below. In this model, we eliminate E_d as the objective function (13.8) and from the resource constraints (13.9a)–(13.9c) and introduce θ_d as the new objective function (now to be maximized) and into the performance enhancement constraints (13.10a)–(13.10d). The parameter θ_d will now be greater than or equal to one, and it is called the *inverse efficiency* of DMU d.

Max θ_d	(13.8)	
subject to		
$\sum_{j=1}^{n} \lambda_j X_{1j} \leq X_{1d}$	(13.9a)	FTE teachers
$\sum_{j=1}^{n} \lambda_j X_{2j} \leq X_{2d}$	(13.9b)	FTE teacher support
$\sum_{j=1}^{n} \lambda_j X_{3j} \leq X_{3d}$	(13.9c)	Building administration and professional staff
$\sum_{j=1}^{n} \lambda_j Y_{1j} \geq \theta_d Y_{1d}$	(13.10a)	Secondary level English (%)
$\sum_{j=1}^{n} \lambda_j Y_{2j} \geq \theta_d Y_{2d}$	(13.10b)	Secondary level math (%)

(continued)

$\sum_{j=1}^{n} \lambda_j Y_{3j} \geq \theta_d Y_{3d}$	(13.10c)	Grade 8 science (%)
$\sum_{j=1}^{n} \lambda_j Y_{4j} \geq \theta_d Y_{4d}$	(13.10d)	Graduation rate (%)
$\sum_{j=1}^{n} \lambda_j S_{1j} \geq S_{1d}$	(13.11a)	Number of elementary school students
$\sum_{j=1}^{n} \lambda_j S_{2j} \geq S_{2d}$	(13.11b)	Number of secondary school students
$\sum_{j=1}^{n} \lambda_j S_{3j} \geq S_{3d}$	(13.11c)	Percentage of students with free or reduced price lunch
$\sum_{j=1}^{n} \lambda_j S_{4j} \geq S_{4d}$	(13.11d)	Percentage of students with limited English proficiency
$\sum_{j=1}^{n} \lambda_j S_{5j} \leq S_{5d}$	(13.11e)	School district's combined wealth ratio
$\sum_{j=1}^{n} \lambda_j = 1$	(13.12)	Variable returns to scale
$\lambda_j \geq 0 \, for \, j = 1, 2, \ldots, 624$	(13.13)	Nonnegativity
$\theta_d \geq 0$	(13.14)	Nonnegativity

The Mixed DEA Model

The mixed DEA model with variable returns to scale, for DMU d, $d = 1, 2, \ldots,$ 624, is below. In this model, we keep both E_d and θ_d in the constraints and we may now choose to either minimize θ_d or maximize θ_d. We introduce a new constraint (13.20) that ensures balance between the goals of reducing resources and enhancing performance.

Min E_d *or Max* θ_d	(13.15)	
subject to		
$\sum_{j=1}^{n} \lambda_j X_{1j} \leq E_d X_{1d}$	(13.16a)	FTE teachers
$\sum_{j=1}^{n} \lambda_j X_{2j} \leq E_d X_{2d}$	(13.16b)	FTE teacher support
$\sum_{j=1}^{n} \lambda_j X_{3j} \leq E_d X_{3d}$	(13.16c)	Building administration and professional staff
$\sum_{j=1}^{n} \lambda_j Y_{1j} \geq \theta_d Y_{1d}$	(13.17a)	Secondary level English (%)

(continued)

$\sum_{j=1}^{n} \lambda_j Y_{2j} \geq \theta_d Y_{2d}$	(13.17b)	Secondary level math (%)
$\sum_{j=1}^{n} \lambda_j Y_{3j} \geq \theta_d Y_{3d}$	(13.17c)	Grade 8 science (%)
$\sum_{j=1}^{n} \lambda_j Y_{4j} \geq \theta_d Y_{4d}$	(13.17d)	Graduation rate (%)
$\sum_{j=1}^{n} \lambda_j S_{1j} \geq S_{1d}$	(13.18a)	Number of elementary school students
$\sum_{j=1}^{n} \lambda_j S_{2j} \geq S_{2d}$	(13.18b)	Number of secondary school students
$\sum_{j=1}^{n} \lambda_j S_{3j} \geq S_{3d}$	(13.18c)	Percentage of students with free or reduced price lunch
$\sum_{j=1}^{n} \lambda_j S_{4j} \geq S_{4d}$	(13.18d)	Percentage of students with limited English proficiency
$\sum_{j=1}^{n} \lambda_j S_{5j} \leq S_{5d}$	(13.18e)	School district's combined wealth ratio
$\sum_{j=1}^{n} \lambda_j = 1$	(13.19)	Variable returns to scale
$E_d + \theta_d = 2$	(13.20)	Balance resource reduction and performance enhancement
$\lambda_j \geq 0 \, for \, j = 1, \, 2, \, \ldots, \, 624$	(13.21)	Nonnegativity
$E_d, \, \theta_d \geq 0$	(13.22)	Nonnegativity

References

Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision making units. Eur J Oper Res 2:429–444

Charnes A, Cooper WW, Rhodes E (1979) Measuring the efficiency of decision making units: short communication. Eur J Oper Res 3:339

Charnes A, Cooper WW, Rhodes E (1981) Evaluating program and managerial efficiency: an application of data envelopment analysis to program follow through. Manag Sci 27:668–697

Cook WD, Tone K, Zhu J (2014) Data envelopment analysis: prior to choosing a model. Omega 44:1–4

Cooper WW, Seiford LM, Tone K (1999) Data envelopment analysis: a comprehensive text with models, applications, references and DEA-Solver software. Kluwer, Boston

Cooper WW, Seiford LM, Zhu J (2011) Handbook on data envelopment analysis. Springer, New York

Emrouznejad A, Parker BR, Tavares G (2008) Evaluation of research in efficiency and productivity: a survey and analysis of the first 30 years of scholarly literature in DEA. Socioecon Plann Sci 42:151–157

Farrell MJ (1957) The measurement of productive efficiency. J R Stat Soc Ser A 120(3):253–290

Liu JS, Lu LYY, Lu WM, Lin BJY (2013) Data envelopment analysis 1978–2010: a citation-based literature survey. Omega 41:3–15
New York State Department of Education (n.d.) Guide to reorganization of school districts. http://www.p12.nysed.gov/mgtserv/sch_dist_org/GuideToReorganizationOfSchoolDistricts.htm. Accessed 17 Mar 2014
Sexton TR (1986) The methodology of data envelopment analysis. In: Silkman RH (ed) Measuring efficiency: an assessment of data envelopment analysis New Directions for Program Evaluation No. 32. Jossey-Bass, San Francisco, pp 7–29
Sexton TR, Silkman RH, Hogan A (1986) Data envelopment analysis: critique and extensions. In: Silkman RH (ed) Measuring efficiency: an assessment of data envelopment analysis New Directions for Program Evaluation No. 32. Jossey-Bass, San Francisco, pp 73–105
U.S. Census Bureau (2011) Public education finances. http://www2.census.gov/govs/school/11f33pub.pdf. Accessed 17 Mar 2014
U.S. Government Spending (n.d.) New York State spending. http://www.usgovernmentspending.com/New_York_state_spending.html. Accessed 17 Mar 2014

Chapter 14
Assessing Efficiency and Effectiveness in Marketing: Applications of Data Envelopment Analysis—Prelude to Chapters 15 and 16

Gopalkrishnan R. Iyer and Dhruv Grewal

Abstract Academic theory and marketing practice have both grappled with the problem of assessing marketing performance, either in terms of effectiveness of marketing strategies and/or the efficiency of marketing operations. Data Envelopment Analysis (DEA) is an important tool for the assessment of marketing efficiency, or the relation between marketing inputs and outputs, as well as marketing effectiveness, or the impact of marketing actions on specific performance results. The following pages provide an introduction to various challenges in the assessment of marketing performance followed by two articles that demonstrate the application of DEA in two different contexts. In the retail context, DEA is used to highlight the importance of considering regional factors when evaluating the efficiency of individual stores within the same retail chain. In the global marketing context, DEA is used to address the fact that most marketing actions involve the creation of marketing assets that in turn impact performance.

Keywords DEA in marketing • Marketing performance assessment • Retail store efficiency • Subsidiary performance evaluation

Marketing performance assessments have been the subject of intense scrutiny for several decades, with renewed interest in recent years (Gauri 2013; Morgan et al. 2002; Rust et al. 2004; Sheth and Sisodia 2002; Vorhies and Morgan 2003). In particular, attention has centered on the efficiency of marketing actions

---

G.R. Iyer
Center for Services Marketing and Management, Florida Atlantic University, 777 Glades Road, Boca Raton, FL 33431, USA
e-mail: giyer@fau.edu

D. Grewal (✉)
Babson College, 213 Malloy Hall, Babson Park, MA 02457, USA
e-mail: dgrewal@babson.edu

J. Zhu (ed.), *Data Envelopment Analysis*, International Series in Operations Research & Management Science 238, DOI 10.1007/978-1-4899-7684-0_14

(e.g., Bonoma and Clark 1988) and the effectiveness of marketing strategies (e.g., Dunn et al. 1994). Some researchers also emphasize the analytical processes involved in assessing marketing performance (e.g., Laréché and Srinivasan 1982).

Central to any efficiency-based marketing performance assessments are two distinct but complementary elements. First, an assessment of marketing productivity addresses the effective management of the marketing mix and the efficiency of marketing spending (Thomas 1984). Second, the extent to which marketing activities contribute to the firm's bottom line signals the impact of marketing on overall firm performance, measured by financial results (Day and Fahey 1988; Srivastava et al. 1998).

Early approaches to measuring marketing productivity focused more on distribution and logistics, rather than the full span of the marketing function (Sheth and Sisodia 2002). In turn, marketing productivity measures tend to consider the return on marketing investments, at the mass market or segment level. Costs and expenditures can often be measured precisely, but considerations of what to include in measures of marketing performance have long been a bone of contention for academics and practitioners. For example, measures of marketing performance might include market share, sales generated, and customer account data—which also are the most readily available and objective outcomes of marketing activities. However, a view of marketing as an expense rather than an investment often resulted in choices that sacrificed key elements necessary for long-term marketing success, such as customer satisfaction, when the related activities were deemed too costly (Sheth and Sisodia 2002). Other, more recent measures of marketing performance include brand equity (Keller 1993), customer retention (Gummeson 1999), customer equity (Blattberg and Deighton 1996; Rust et al. 2004), customer lifetime value (Berger and Nasr 1998; Kumar and Shah 2009), and quality (Rust et al. 2002).

In addition, the challenge of marketing performance assessments at the top management level has led to an emphasis on marketing's impact on firm performance as a measure of marketing success. Thus, another set of approaches attempts to link marketing investments and/or marketing performance measures to firm performance, with the recognition that marketing performance measures such as brand equity and customer loyalty tend to be intangible outputs (Srinivasan and Hanssens 2009). Prior research cites the links between marketing mix investments and some measures of firm value. For example, researchers have studied the impacts on firm value of investments in advertising (Joshi and Hanssens 2009), sales promotions (Pauwels et al. 2004), distribution channels (Geysens et al. 2002), or product innovations (Srinivasan et al. 2009), among others. Marketing outputs also can be linked to firm value; in their review, Srinivasan and Hanssens (2009) highlight studies that focus on the firm value impacts of brand equity (e.g., Madden et al. 2006), customer satisfaction (e.g., Fornell et al. 2006), customer equity (e.g., Gupta et al. 2004), and product quality (e.g., Tellis and Johnson 2007).

Yet another approach to performance assessment links marketing actions and investments to the creation of market-based outputs, which then contribute to firm performance. Thus, marketing outputs are key intermediate outputs of the links between the marketing function and firm performance. The development and leveraging of market-based assets and their consequent impact on shareholder value are key to these approaches (Day and Fahey 1988; Srivastava et al. 1998). A core thrust is the goal of linking the performance of marketing activities to objective measures that are supra-marketing measures. They involve inclusion of traditional financial measures of firm performance, such as net present value of cash flows and shareholder value, together with marketing measures, such as sales volume, market share, and gross margin (Day and Fahey 1988; Srivastava et al. 1998). The performance output measures are broad, firm-level financial measures; the key drivers or inputs of performance instead involve attention to market-based assets and business processes that are within the marketing domain (Srivastava et al. 1998, 1999).

Effectiveness-based approaches to marketing performance start with the assumption that measuring marketing productivity provides only one side of the story. Several aspects of a marketing strategy cannot be objectively assessed or numerically measured, including the impact of marketing on the achievement of business and organizational goals (Walker and Ruekert 1987). The crux of these approaches is the extent of fit between the core marketing strategy pursued and the marketing organization (McKee et al. 1989; Walker and Ruekert 1987). The key driver of marketing performance in this context is thus the fit between a specific strategy type and the structural and task characteristics of the marketing organization (Slater and Narver 1993; Walker and Ruekert 1987; Workman et al. 1998). The congruence between the marketing strategy and marketing organization is important not only for assessing marketing effectiveness but also for competitive reasons (Bharadwaj et al. 1993; Day 1994). The effectiveness of the fit is not visible to competitors, which makes it difficult to imitate (Bharadwaj et al. 1993). Accordingly, the fit between strategy and organization appears critical to marketing performance (Moorman and Rust 1999; Workman et al. 1998).

Viewed according to the performance assessment framework, efficiency-based approaches focus on tangible measures of both inputs and outputs (Bonoma and Clark 1988), whereas effectiveness-based approaches focus on intangible measures of marketing strategy (e.g., ideal strategic types) and marketing organization (Walker and Ruekert 1987). Both approaches have been extended to include measures of organizational performance. In the case of efficiency-based approaches, they include measures of the firm's financial performance (Srivastava et al. 1998), but in the case of effectiveness-based measures, they include broad measures of organizational performance (Conant et al. 1990). At another level, focusing on the intangibility of outputs has resulted in other measures of marketing performance, such as brand equity and customer value.

At the heart of all approaches to broadening marketing performance assessment lie two academic and practical pressures: (1) the assessment of the impacts of marketing on the overall organization and (2) the assessment of intermediate processes through which marketing contributes to firm performance. These objectives have led to various discourses, including a focus on the financial performance implications of marketing (Srivastava et al. 1998); the merger of efficiency and effectiveness dimensions of marketing performance, in the form of "effective efficiency" (Sheth and Sisodia 2002); measures of both efficiency and effectiveness (Vorhies and Morgan 2003); and the development of tangible measures of various aspects of marketing inputs or outputs (Berger and Nasr 1998; Rust et al. 2004).

Against this lack of consensus about the appropriate inputs, outputs, or processes for marketing performance, it becomes clear that marketing performance assessment cannot be considered straightforward or easy. Chapters 15 and 16 provide examples of assessments at the retail level or the global level.

Retail Level: In Grewal et al. (1999), "Planning Merchandising Decisions to Account for Regional and Product Assortment Differences," DEA serves to assess the efficiency of 59 stores of a major *Fortune* 500 multinational retailing chain. Multiple inputs (store operating expenses, square footage of each store, inventory) lead to the key outputs of sales volume in units for sales of products A and B. A regional factor serves as the control variable. Chapter 15 (Grewal et al. 1999) highlights the importance of incorporating regional differences in the analysis and potential adjustments to target sales on the basis of assessments of how the best practice stores are performing.

Marketing Performance Assessments at a Global Level: Empirical assessments of global marketing performance are relatively sparse. When they exist, they tend to focus on specific aspect(s) of marketing strategy and firm performance. Because a global firm learns from its operations in various countries, assessments of global marketing performance must not only compare the relative performance of each subsidiary but also focus on the processes that lead to superior performance. Grewal et al. (2009), in Chap. 16 acknowledge that in international markets, the firm first must create certain key capabilities before it can achieve success in each market in which it operates (Luo 1999). These capability-building efforts lead to the development of critical market assets (Srivastava et al. 1998) that can be leveraged to enhance marketing performance. Strategic moves by the firm and its subsidiaries to develop national markets leads to market creation; harnessing national markets in terms of their observable performance is termed market yield.

Specifically, the development of critical, intangible, and rent-generating assets by subsidiaries is a first step toward enhancing marketing performance, not only for the subsidiary but for the firm as a whole. Knowledge is one such critical asset, though other market-based assets exist too. A focus on the intermediate development of critical resources makes it imperative that marketing performance goes beyond simply the relationship between inputs and outputs. It must involve the development of enduring market-based assets that, when harnessed, contribute to superior performance. Therefore, marketing performance assessments must focus on process aspects, as well as inputs and outputs.

The national environment of each subsidiary also influences global marketing operations and performance. They variously might be sources of competitive advantage creation, locus of location-specific advantages, wellsprings for different factor costs or margins, barriers to transfers of extant knowledge, or schools for developing experiential knowledge. Any performance assessment of global operations must account for the environment. In general, the environment moderates the development of market-based assets, such as knowledge, experience, and firm-specific competitive advantages, as well as the extent to which those market-based assets can be harnessed to produce superior performance.

Chapter 16 (Grewal et al. 2009) use DEA to assess marketing performance using multiple inputs and outputs. Moreover, they apply DEA to identify subsidiaries that are best in terms of either or both market creation and market yield. Thus, DEA can be applied not just for performance assessments but also to identify subsidiaries that provide insights into best practices, to be developed and transmitted across the global organization.

References

Berger PD, Nasr NI (1998) Customer lifetime value: marketing models and applications. J Interact Mar 12:17–30

Bharadwaj SG, Rajan Varadajaran P, Fahy J (1993) Sustainable competitive advantage in service industries: a conceptual model and research propositions. J Mark 57:83–99

Blattberg RC, Deighton J (1996) Manage marketing by the customer equity test. Harv Bus Rev 74:136–144

Bonoma TV, Clark BH (1988) Marketing performance assessment. Harvard Business School Press, Boston

Conant JS, Mokwa MP, Varadarajan PR (1990) Strategic types, distinctive marketing competencies and organizational performance: a multiple measures-based study. Strat Manage J 11 (5):365–383

Day G (1994) The capabilities of market-driven organizations. J Mark 58:37–52

Day G, Fahey L (1988) Valuing market strategies. J Mark 52:45–57

Dunn MG, Norburn D, Birley S (1994) The impact of organizational values, goals, and climate on marketing effectiveness. J Bus Res 30(2):131–141

Fornell C, Mithas S, Morgeson F, Krishnan MS (2006) Customer satisfaction and stock prices: high returns, Low risk. J Mark 70:3–14

Gauri DK (2013) Benchmarking retail productivity considering retail pricing and format strategy. J Retail 89(1):1–14

Geysens I, Gielens K, Dekimpe MG (2002) The market valuation of Internet channel additions. J Mark 66(2):102–119

Grewal D, Levy M, Mehrotra A, Sharma A (1999) Planning merchandising decisions to account for regional and product assortment differences. J Retail 75(3):405–424

Grewal D, Iyer GR, Kamakura W, Mehrotra A, Sharma A (2009) Evaluation of subsidiary marketing performance: combining process and outcome performance metrics. J Acad Mark Sci 37(2):117–129

Gummeson E (1999) Total relationship marketing—rethinking marketing management: from 4P's to 30 R's. Butterworth-Heineman, Oxford

Gupta S, Lehmann DR, Stuart JA (2004) Valuing customers. J Mark Res 41(1):7–18

Joshi AM, Hanssens DM (2009) Movie advertising and the stock market valuation of studios: a case of 'great expectations'? Mark Sci 28(2):239–250

Keller KL (1993) Conceptualizing, measuring, and managing customer-based brand equity. J Mark 57:1–22

Kumar V, Shah D (2009) Expanding the role of marketing: from customer equity to market capitalization. J Mark 73:119–136

Laréché J-C, Srinivasan V (1982) STRATPORT: a model for the evaluation and formulation of business portfolio strategies. Manage Sci 28(9):979–1001

Luo Y (1999) Entry and cooperative strategies in international business expansion. Quorum Books, Westport

Madden TJ, Fehle F, Fournier S (2006) Brands matter: an empirical demonstration of the creation of shareholder value through branding. J Acad Mark Sci 34(2):224–235
McKee DO, Rajan Varadarajan P, Pride WM (1989) Strategic adaptability and firm performance: a market-contingent perspective. J Mark 53:21–35
Moorman C, Rust RT (1999) The role of marketing. J Mark 63(Special Issue):180–197
Morgan NA, Clark BH, Gooner R (2002) Marketing productivity, marketing audits, and systems for marketing performance assessment: integrating multiple perspectives. J Bus Res 55 (5):363–375
Pauwels KH, Silva-Risso JM, Srinivasan S, Hanssens DM (2004) New products, sales promotions and firm value, with application to the automobile industry. J Mark 68:142–156
Rust RT, Moorman C, Dickson PR (2002) Getting return on quality: revenue expansion, cost reduction, or both. J Mark 66:7–24
Rust RT, Moorman C, Dickson PR, Lemon KN, Zeithaml VA (2004) Return on marketing: using customer equity to focus marketing strategy. J Mark 68:109–127
Sheth JN, Sisodia RS (2002) Marketing productivity issues and analysis. J Bus Res 55(5):349–362
Slater SF, Narver JC (1993) Product-market strategy and performance: an analysis of the miles and snow strategic types. Eur J Mark 27(10):33–51
Srinivasan S, Hanssens DM (2009) Marketing and firm value: metrics, methods, findings, and future directions. J Mark Res 46:293–312
Srinivasan S, Pauwels K, Silva-Risso J, Hanssens DM (2009) Product innovations, advertising, and stock returns. J Mark 73(1):24–43
Srivastava RK, Shervani TA, Fahey L (1998) Market-based assets and shareholder value: a framework for analysis. J Mark 62:2–18
Srivastava RK, Shervani TA, Fahey L (1999) Marketing, business processes, and shareholder value: an organizationally embedded view of marketing activities and the discipline of marketing. J Mark 63(Special Issue):168–179
Tellis GJ, Johnson J (2007) The value of quality. Mark Sci 26(6):758–773
Thomas MJ (1984) The meaning of marketing productivity analysis. Mark Intell Plan 2(2):13–28
Vorhies DW, Morgan NA (2003) A configuration theory assessment of marketing organization Fit with business strategy and its relationship with marketing performance. J Mark 67:100–115
Walker OC, Ruekert RW (1987) Marketing's role in the implementation of business strategies: a critical review and conceptual framework. J Mark 51:15–33
Workman JP, Homburg C, Gruner K (1998) Marketing organization: an integrative framework of dimensions and determinants. J Mark 62:21–41

Chapter 15
Planning Merchandising Decisions to Account for Regional and Product Assortment Differences

Dhruv Grewal, Michael Levy, Anuj Mehrotra, and Arun Sharma

Abstract The last decade has fundamentally changed the face of retailing. The genesis has been increased customer fragmentation, enabling technologies such as the internet and increased competition. In this era of "hypercompetition," retailers need to have a better understanding of the performance of individual stores so they can more accurately plan their merchandise assortments and set more realistic merchandising goals. In this paper, we determine the performance of retail outlets relative to the "best practice" set of outlets and demonstrate the importance of accommodating both regional and assortment differences. We empirically assess the performance of stores from a major Fortune 500 multinational retailing chain. Multiple inputs and outputs from 59 stores in three regions were used to determine sales goals for two different product categories. The results of three alternative models suggest that incorporating both assortment and regional differences significantly affects both performance and predicted sales volume estimates. Implications and avenues for future research are discussed.

Keywords Retailing • Merchandise assortment • Store efficiency • Assortment

The retail environment has become much more competitive over the past few decades. The growth of national specialty store chains (e.g., The Gap, Crate and Barrel) and category killers (e.g., Toys 'R' Us, Sports Authority) have significantly altered the retail landscape. These retailers have tended to take over their respective categories and consequently decimated many smaller, less efficient retailers.

This chapter is reproduced from, Grewal, Dhruv, Michael Levy, Anuj Mehrotra and Arun Sharma (1999), "Planning Merchandising Decisions to Account for Regional and Product Assortment Differences," *Journal of Retailing*, 75 (3), 405–424, with permission from New York University, Stern School of Business

D. Grewal (✉) • M. Levy
Babson College, 213 Malloy Hall, Babson Park, MA 02457, USA
e-mail: dgrewal@babson.edu

A. Mehrotra • A. Sharma
University of Miami, Coral Gables, FL, USA

J. Zhu (ed.), *Data Envelopment Analysis*, International Series in Operations Research & Management Science 238, DOI 10.1007/978-1-4899-7684-0_15

There has also been an increase in customer fragmentation as smaller and smaller groups of customers demand products and services tailored to their individual needs (Bessen 1993; Blattberg and Deighton 1991; Kahn and McAlister 1997). Retailers are responding to these customer demands by embracing new technologies such as data base marketing and mass customization (Gilmore and Joseph Pine 1997). For example, made-to-measure Levi's are now available in some of their stores. Similarly, Custom Foot in Westport Connecticut can deliver a custom-made shoe from Italy in 2 weeks at mass-produced prices.

Another key driver of change in the structure of retailing is electronic commerce. For example, Dell and Gateway computers are each selling more than three million dollars every day on the Internet, whereas computer sales through traditional retailers have had flat to declining revenues (*Business Week*, March 23, 1998, pp. 28–31).

All of these change drivers (e.g., category specialists, customer fragmentation, and technology) suggest that store-based retail chains must maintain cutting edge information systems to compete. In particular, they need to be able to accurately measure merchandise performance, predict the sale of merchandise categories, and identify and correct problems at the individual store level. For example, an automobile dealer may be a highly rated franchisee because its total sales of automobiles and light trucks are high. However, closer analysis might indicate that the proportion of light trucks sold is well below the norm, compared with similar dealers with similar locational properties. This type of performance information, along with a prediction of what the sale of light trucks should be, would provide a signal of potential problems. Further investigation might determine, for instance, that the sales of light trucks were low because of a poor inventory position, ineffective sales training, or poor promotions.

The previous example highlights the need to better understand the operations of individual stores so that their performance can be maximized. Unfortunately retailers typically rely on aggregate measures, such as sales per square foot, gross margin, inventory turnover, and GMROI, to plan and evaluate the performance of individual stores and departments within those stores. These measures are used to compare the performance of different stores, departments, as well the managers and buyers who run them. Unfortunately, as the car dealer example indicated, they can provide false signals for both planning and evaluating merchandising performance. Instead, we suggest that the planning and evaluation of merchandising activities be performed at a more disaggregate level using Data Envelopment Analysis (DEA). DEA allows managers to plan and evaluate the performance of similar operating entities. In essence apples are compared with other apples, rather than oranges.

The objective of this paper, therefore, is to demonstrate how to measure better the efficiency of a retail chain by dissagregating sales to consider certain characteristics that may effect a category's or a store's performance. In particular, we examine the effects of the disaggregation of two factors: regional and assortment differences. It is important for retailers to have a complete understanding of the efficiency of each outlet so they can distinguish between excellent and mediocre performers. Once the parameters of excellence are known, the mediocre stores can be modeled after the excellent ones. Also, because store managers' evaluations are based on the performance of their stores, it is important that their performance be evaluated in a fair and equitable fashion. Retailers also need to understand exactly why a store is performing the way that it is. For instance, are the successes or failures a result of managerial action, or an artifact of some environmental factor? In spite of the importance of the topic, the issue has not been extensively examined in the literature (see Kamakura et al. 1996).

There are three benefits to the proposed method. First, stores are evaluated against comparable "best practice" stores rather than the "average performers" traditionally used in mean-based analyses such as multiple regression. Not only is this type of evaluation more fair and accurate, but it also enables managers to determine how far they need to go to become a "best practice" store based on slack analyses. Second, performance is disaggregated at the category level, thus avoiding the potentially misleading and even inaccurate performance benchmarks inherent in using overall sales. Finally, regional differences are controlled because these differences can play a major role in determining the performance of a store.

We first examine the underpinnings of our approach to assortment planning. This discussion is followed by an examination of the concept of efficiency, and how retailers can utilize information on store efficiency in their merchandise planning process. Then, we illustrate the use of DEA for merchandise planning purposes. Specifically, the results of three analyses that test and assess the effects of assortment and regional differences on efficiency estimates and sales projections are reported. Finally, implications and avenues for future research are discussed.

15.1 Assortment and Regional Planning, and Store Performance

The best retailers have learned to adjust their assortment by region to better meet the needs of customers. For example, Burdines, the Florida-based division of Federated Department Stores, carries a somewhat different assortment of goods than Macy's (although they are also owned by Federated). Because of the assortment differences, it would not be fair to compare the performance of a Burdines store with a Macy's store in Orlando, Florida. This situation is further complicated if regional effects are taken into consideration. Macy's roots are in New York, whereas Burdines is a Florida chain. Therefore, even if the stores were to carry exactly the same merchandise, the patronage of these stores would be somewhat dependent upon the number of "New York Transplants" shopping in Florida to the "Native Floridians." Therefore, the Macy's store in Orlando, Florida should not be directly compared with either a Burdines store in Orlando or a Macy's store in New York. The correct comparison would be another Macy's store in Florida.

With this introduction, we review the literature on assortment and regional planning for retailers. Retailers have traditionally varied their assortments to appeal to specific customer groups, provide variety to customers, and to cross-sell products and services. The merchandise assortment itself has become an effective method to attract and retain core customers. Harley Davidson has effectively used this strategy by combining lifestyle accessories in combination with their motorcycles.

Range Rover has also successfully used this formula by redesigning their dealerships to teach customers how to drive Range Rovers as well as sell merchandise. Second, merchandise assortment strategies, such as scrambled merchandise, provide variety to customers and appeal to variety-seekers (Kahn 1991; Kahn and Lehmann 1991; Kahn and McAlister 1997). Finally, bundling products and services facilitates cross selling to larger segments (Yadav and Monroe 1993). For example, Best Buy sells a bundle consisting of a computer and an extended service contract to a segment that typically only bought computers and did not consider their service needs.

Retail chains are developing store formats that integrate regional differences to better meet customer needs. They realize, for instance, that more petite sizes are needed in areas where there is a high Hispanic or Asian population. Also, wider assortments of apparel should be found in smaller towns because there are usually fewer apparel outlets for customers than large towns offer. How do some stores fine-tune their assortments? Target's micromarketing efforts are based on age, climate, small-town community, and African American, Hispanic, or Asian heritage (Solomon 1994). These stores and others have found that having distinctive assortments tailored to specific customer groups is a viable method of developing a distinctive strategic competitive advantage. Unfortunately, these regional differences complicate the assortment planning process, and the methods used to achieve efficient regional assortments are not well defined in the literature.

15.2 Retail Productivity and Efficiency

Understanding and measuring the productivity and efficiency of retailers have been important issues in retailing research (e.g., Bucklin 1978; Donthu and Yoo 1998; Ingene 1982, 1984; Ratchford and Stoops 1988; Ratchford and Brown 1985). Past research has used and suggested the use of various measures and methods to assess retail efficiency. The majority of measures of outlet efficiency are input-output ratios, such as sales per square foot or sales per employee (Kamakura et al. 1996).

These traditional methods are problematic when a retail chain has multiple goals. Take, for example, a typical computer store that sells both products (e.g., computers, printers, etc.) and services (e.g., repairs, add-ons, etc.). They want to maximize the sales of both these outputs. Traditional methods would sum these two outputs, but would be unable to identify the optimal level of the individual outputs. Similarly, stores have both personnel and merchandise inputs. Traditional efficiency analysis methods would combine these two inputs into a single expenditure measure. These two inputs should, however, remain separate because they are not substitutable. Otherwise, the store might have great merchandise, but poor sales help, or vice versa. Using a combined input measure, management would not be able to delimit the problem.

Kamakura et al. (1996) and Donthu and Yoo (1998) identified a number of problems with traditional input/output measures. We highlight the following problems that are relevant for the purposes of this study:

1. Most retailing situations have multiple inputs and outputs that are not addressed in traditional analysis (Kamakura et al. 1996).
2. Most measures are sales management oriented (e.g., sales per hour) rather than measures of total retail productivity (Donthu and Yoo 1998).
3. Output is defined as a supply-side concept. The market conditions are normally not included in the analysis (Donthu and Yoo 1998; Ratchford and Stoops 1988).
4. Research that has used standard regression analysis used average store performance rather than the best performers Chebat et al. (1994) and (Donthu and Yoo 1998).
5. Very few studies have examined the context of the efficiency of different outlets of a firm (Donthu and Yoo 1998; Kamakura et al. 1996).

This study examined the efficiency of outlets by using an efficiency frontier methodology called DEA that has started to be adopted in various marketing settings. Kamakura et al. (1988) examined brand efficiencies based on attribute data of brands and compared efficiencies with retail prices. Mahajan (1991) examined the efficiency of salespeople. Similarly, Boles et al. (1995) and Parsons (1990) used DEA to evaluate the performance and rank salespeople while using multiple input and outputs. Kamakura et al. (1996) evaluated bank-outlet performance and determined cost functions. Donthu and Yoo (1998) evaluated retail store performance and compared results with regression analysis. Murthi et al. (1996) calculated the efficiency of firms relative to that of competing stores in a mature consumer goods market. Although some of these studies were in a retail setting, none have addressed the issues of calculating performance while accommodating regional and assortment differences.

There are four reasons why this methodology enables retailers to better plan their assortments. First, DEA accommodates multiple inputs and outputs, thus allowing for the disaggregation of total sales volume into individual product categories. This disaggregation process enables managers to better understand the assortment needs of each individual store. Second, the method allows for the inclusion of regional differences in retail outlets. Third, the method allows a comparison between individual stores with the best stores in the chain. Finally, DEA enables firms to predict what the sales of a store would be if it were performing as a best practice store. This sales prediction takes both regional and assortment differences of individual stores into consideration. This "best practice" prediction allows retailers to set more realistic and accurate goals based on the specific store profiles.

Traditional methods of evaluating store performance typically use aggregated sales data. The problem with using aggregated data is that it leads to optimistically large sales target goals that may be unrealistic and unreachable because it ignores a particular store's advantage in selling a particular product category over another. Also, in a traditional aggregate level analysis, some stores are likely to be regarded as successful or efficient stores simply because they are in better regions or locations.

Using DEA, sales can be disaggregated by product category and by region. Stores will be evaluated against “best practice” stores in the same region and with similar assortments. To illustrate how the analysis using disaggregated sales data works, consider three stores that all carry category “A” and “B” and all have identical inputs.

The first store sold seven units of “A” and seven units of “B.” The second store sold 5 units of “A” and 15 units of “B.” The third store sold eight units of “A” and eight units of “B.” Using aggregated sales, the second store is efficient with 20 units sold, and stores one and three are inefficient. With disaggregated data, both the second and third stores are efficient because the second store sold the most of category “B” and the third store sold the most of category “A.” Importantly, if the first store sold only one more each of products “A” and “B,” it too would be efficient. Yet, using aggregated data, the first store would be in third place, and would have to increase aggregate sales by at least six units to make it efficient. The use of aggregated sales data may therefore obfuscate reality and erroneously penalize stores carrying categories “A” and those carrying both “A” and “B.”

15.2.1 Model Development

A fundamental assumption behind DEA analysis is that if a given store-A, is capable of y_A units of outputs (e.g., sales) with x_A inputs, then other stores should also be able to do the same if they were to operate efficiently. Similarly, if store-B is capable of y_B units of outputs with x_B inputs, then other stores should also be capable of the same performance.

DEA also assumes that stores A, B, and others can be combined to form a composite store with composite inputs and composite outputs. Because this composite store may not necessarily exist, it is typically called a virtual store.[1] The heart of the analysis lies in finding the “best” virtual store for each real store. If the virtual store is better than the store being evaluated by either accomplishing more outputs with the same inputs or having the same outputs with less inputs, then the original store is inefficient.

The procedure for finding the best virtual store can be formulated as a linear program. Analyzing the efficiency of N stores is then a set of N linear programming problems. The efficiency frontier defines the maximum combinations of outputs that can be posted for a given set of inputs.

The DEA literature proposes several different types of models that have been developed for a variety of different goals. The description of these models and their differences are beyond the scope of this paper. We restrict our attention to describing the model suggested for calculating (technical) inefficiencies (see, Banker and Morley 1986a, b) of stores in an output formulation: one in which we focus on

[1] Virtual stores is a term used in DEA analysis and does not refer to an Internet-based store.

considering the estimation of the extent to which outputs could be increased without requiring additional inputs. This particular model is appropriate for the evaluation of retail stores because one of the primary objectives of most retail chains is to maximize sales and market share.

Suppose $output_{rj}$, r ϵ {1,. . .,S} and $input_{ij}$, for $i \in I = \{1,...,M\}$ are the observed output and input values for $j = 1,...,N$ stores. The linear programming problem that helps estimate the output technical inefficiency measure for *store x* is as follows:

$$Maximize\, score + \in \left(\sum\nolimits_{r=1}^{s} excess_r + \sum\nolimits_{r=1}^{M} slack_i\right)$$

Subject to :

$$\sum\nolimits_{j=1}^{N} \lambda_j input_{ij} + slack_i = \mathrm{input}_{ix}, i \in \{1, \ldots, M\}$$

$$\sum\nolimits_{j=1}^{N} \lambda_j output_{rj} - excess_r = \mathrm{score}^*\mathrm{output}_{rx}, r \in \{1, \ldots S\}$$

$$\sum\nolimits_{j=1}^{N} \lambda_j = 1$$

$$score, \lambda_i, excess_r, slack_i \geq 0$$

The first constraint set indicates that the weighted sum of inputs of the virtual store is set equal to at most the input of the store under investigation. The slack $slack_i$ can take positive values when the virtual store does not need to use inputs at the same level. The second set of constraints indicates that corresponding to using these levels of inputs, the virtual store is capable of producing outputs at the level score $*$ $output_{rx}$ where $score \geq 1$and ϵ is an infinite by small, positive number.

DEA determines a store to be efficient only when comparisons with other relevant stores do not provide evidence of inefficiency in the use of any input or output. An efficient store has a score of 1, and has $slack_i = 0$ for each input i, i.e., it is on the efficiency frontier. The closer the score is to 1, the more efficient the store is considered to be. In output oriented models, scores are greater than or equal to 1. This is in contrast to input-oriented models, in which scores vary from 0 to 1. The reasons that we use inefficiency scores that are greater or equal to one, is that the scores are an indication of the factor improvement in output of specific decision making units (DMUs) to make them efficient.

For an inefficient store, the adjustments needed in each of its outputs to render it technically efficient is given by:

$$\mathrm{output}'_{rx} = \mathrm{excess}_r + \mathrm{score}^*\mathrm{output}_{rx} \text{ for } r = 1, \ldots, S.$$

Simultaneously, inefficient scores must decrease their input levels by *slacki*. In other words, slack analysis helps retail outlets allocate resources more efficiently and improve their performance. Also, slack analysis enables managers to identify their store's potential.

15.3 Empirical Illustration

The efficiency of 59 retail outlets of a Fortune 500 retailing and manufacturing firm is calculated using DEA. To empirically demonstrate the effect of assortment and regional differences on individual store assessments, "best practice" based performance analysis is calculated on an aggregate basis, an assortment basis, and on a regional assortment basis. The results are then compared with a regression-based analysis.

Confidentiality requirements do not allow for the identification of the products or regions. However, the firm is an automobile parts retail chain. The retail outlets sell two non-substitutable categories (A and B) and the majority of their sales come from these two categories. The retail stores are located in three geographical regions (1, 2, and 3). The locational profiles of stores within a region were similar. The first region, in the Northeast, is characterized by cold winters with snow. The average per-capita income within a 3-mile radius of each store was $25,177. The second region, in the Midwest, has very cold winters with large amounts of snow. The average per-capita income within a 3-mile radius of each store was $26,538. Finally, the third region, on the west coast, has a temperate climate. The average per-capita income within a 3-mile radius of each store was $22,389. Also, because one region is more rural than the others, its residents use a larger proportion of light trucks, compared with automobiles. As a result, the demand for certain auto products, such as batteries, tires, and coolants vary dramatically across regions.

15.3.1 Determination of Inputs and Outputs

Researchers have suggested that to evaluate stores, the input factors should include store specific factors (see Donthu and Yoo 1998). Store specific factors can include those pertaining to the square footage of the store, inventory levels or investment, technology, service levels, number of employees, operating hours, operating expenditures, etc. (Bucklin 1978; Lusch and Serpkenci 1990). The store specific input factors used in this study were selected based on management input and previous literature (e.g., Boles et al. 1995; Kamakura et al. 1996; Lusch and Jaworski 1991). They are:

- *Store Operating Expenses* include all operating costs including salary and benefits, store supplies, and utilities, but exclude inventory costs and other corporate costs such as national advertising;
- *Square Footage* represents the size of the store; and
- *Inventory* represents the level of product availability of a specific product and can be viewed as a proxy of reducing buyers' waiting time.

The factors of store operating expenses and inventory were under the control of management. There was high variance in these inputs and the proportion of maximum/minimum was 6.9 for store operating expenses, 4 for square footage, and 4.4 for inventory.

The key output factor is typically sales volume in either dollars or units (Bucklin 1978; Ratchford and Stoops 1988). The output factor used in this study was sales volume in units. The unit cost and retail prices for all SKUs within a category in this case are similar. Thus, the results would be the same if either unit or dollars sales were used.

The purpose of the first study was to determine the aggregate efficiency of individual stores. The single output was sales volume of both products A and B. The purpose of the second analysis was to disaggregate the sales of the two product categories to specifically take assortment differences into consideration. Therefore, two outputs—unit sales of Product A and Unit sales of Product B were used in the analysis. Finally, the third analysis, included the regional factor as a control input variable.

As suggested earlier, we use DEA to determine the performance of retail outlets. Specifically, we calculate the following:

1. Inefficiency Score that is greater than or equal to 1.
2. Sales goals that is equivalent to projected sales of product(s) for the retail outlets to be efficient. This is calculated by adding excess and the product of the present sales and the inefficiency score.
3. Slack in percentage that is calculated by (Sales goals−Current Sales)/(Current Sales). This figure provided outlets with an indication of the sales increase necessary to become efficient.

15.4 Analyses

Three analyses were performed in this study, each with a different level of data aggregation. Each analysis used three inputs (store operating expenses, square footage and inventory) and one or two outputs (sales volumes of product category A, B, or combined). The first analysis provides a measure of each store's relative efficiency based on aggregated sales data. The second analysis uses the same input variables as the first analysis, but disaggregates the output sales data into the two separate major product categories (A and B) and used sales of each product category as different outputs. We will show that disaggregating the sales data provides a better understanding of the role that product assortment plays in the evaluation of a store's performance.

The third analysis is similar to the second, except the comparison is limited to stores from the same region. Specifically, instead of using all 59 stores in one analysis, three different sets of analyses were performed, one for each region. This further refinement allows for the examination of regional differences. For example, suppose the two regions are the greater Sacramento, California area and the greater New York City area. It would be inherently unfair to aggregate the two areas and compare Sacramento stores against those in New York, because the New York stores may be more efficient due to the greater population concentration.

Controlling for a region can be succinctly accomplished by altering the linear program by inclusion of categorical control variables (see Banker and Morley 1986a). Alternatively, the same results are found if one were to run the analysis for each of the three regions separately. Limiting the analysis to a particular region restricts the linear program even further than in the second analysis. This has the effect of generating an even lower inefficiency score for any given store than was possible in the first two analyses. A lower score means that a store will be closer to "efficient" than would be the case without parceling the data into regions and disaggregating by product category. Also, restricting the analysis into regions will result in more conservative estimated sales goals.

15.5 Results

The inefficiency scores as well as the projected sales volumes at the efficient levels for the 59 retail stores from the three geographic regions are presented in Table 15.1.

15.5.1 Role of Dissagreggating Overall Sales to Account for Product Assortment

The comparison of the results of Analysis 1 versus Analysis 2 provides a test of the role of disaggregating sales volume using multiple outputs as opposed to a single summated output. The results reported in Table 15.1 indicate that 10 stores are classified as efficient in Analysis 1, whereas the number increases to 14 in Analysis 2. Thus by considering the efficiency of the two product categories separately, more stores are rated as being efficient. Furthermore, the overall average inefficiency declines from 1.39 in Analysis 1–1.29 in Analysis 2. The decline in efficiency is significant (Table 15.2: $t = 4.96$, $p < .001$). Additionally, the projected sales also decline ($t = 5.29$, $p < .001$).

15.5.2 Role of Regional Differences

The comparison of the results for Analysis 2 versus Analysis 3 provides a test of the role of regional differences (i.e., using three analyses to assess the role of the regions versus a single analysis where the data are pooled). The results reported in Table 15.1 indicate that 14 stores are classified as efficient in Analysis 2, whereas

Table 15.1 Efficiency and total target sales estimates

	Experiment 1		Experiment 2				Experiment 3				
Store	Inefficiency	Total target sales	Inefficiency	Category A target sales	Category B target sales	Total Target Sales	Region	Inefficiency	Category A target sales	Category B target sales	Total target sales
1	1.09	16,687	1.09	1239	15,329	16,568	1	1.03	1254	14,546	15,800
2	1.39	18,085	1.3	1625	15,222	16,847	1	1	1253	11,740	12,993
3	1.48	19,378	1.44	1973	16,880	18,853	1	1	1366	11,685	13,051
4	1.15	22,417	1.15	1493	20,878	22,371	1	1	1298	18,151	19,449
5	1.46	19,682	1.46	1547	18,097	19,644	1	1.11	1254	13,767	15,021
6	1.81	19,907	1.8	1430	18,380	19,810	1	1.42	1254	14,529	15,783
7	1.2	23,367	1.15	2164	20,307	22,471	1	1.07	2008	18,841	20,849
8	1.05	24,861	1	2644	21,019	23,663	1	1	2644	21,019	23,663
9	1.16	24,861	1	2213	19,262	21,475	1	1	2213	19,262	21,475
10	1.15	22,887	1.06	2280	18,942	21,222	1	1	2142	17,799	19,941
11	1.21	22,890	1.17	2109	19,977	22,086	1	1.07	1926	18,241	20,167
12	1	15,962	1	1254	14,708	15,962	1	1	1254	14,708	15,962
13	1.35	17,805	1.05	1484	13,522	15,006	1	1	1415	11,737	13,152
14	1.87	17,645	1.55	1619	12,994	14,613	1	1	1043	8371	9414
15	1.21	24,817	1.16	2176	21,697	23,873	2	1.01	1890	18,850	20,740
16	1.46	18,987	1.45	1459	17,439	18,898	2	1.19	1202	14,372	15,574
17	1	17,965	1	1490	16,475	17,965	2	1	1490	16,475	17,965
18	1.19	19,548	1.12	1583	16,831	18,414	2	1.02	1440	15,314	16,754
19	1.14	24,861	1.12	1826	22,728	24,554	2	1.05	1706	21,230	22,936
20	1.81	14,761	1.02	1323	10,649	11,972	2	1	1293	6851	8144
21	1	13,840	1	951	12,889	13,840	2	1	951	12,889	13,840
22	1.42	17,210	1.42	1525	15,648	17,173	2	1.05	1130	11,590	12,720
23	1	24,861	1	1544	23,317	24,861	2	1	1544	23,317	24,861

(continued)

Table 15.1 (continued)

Store	Experiment 1		Experiment 2				Experiment 3				
	Inefficiency	Total target sales	Inefficiency	Category A target sales	Category B target sales	Total Target Sales	Region	Inefficiency	Category A target sales	Category B target sales	Total target sales
24	1.71	15,100	1.56	1597	12,189	13,786	2	1	1021	7793	8814
25	1.6	15,120	1.6	1157	13,902	15,059	2	1.21	874	10,501	11,375
26	1.28	19,157	1.28	1649	17,474	19,123	2	1.07	1379	14,615	15,994
27	1.44	18,714	1.44	1663	17,004	18,667	2	1.18	1362	13,925	15,287
28	1	10,592	1	521	10,071	10,592	2	1	521	10,071	10,592
29	1.54	21,325	1.39	2323	17,179	19,502	2	1.2	2017	14,638	16,655
30	1.55	20,458	1	2085	11,134	13,219	2	1	2085	11,134	13,219
31	1.24	19,272	1.23	1740	17,360	19,100	2	1.04	1480	14,773	16,253
32	1.14	21,503	1.10	2165	18,482	20,647	2	1	1974	16,855	18,829
33	1.25	18,770	1.04	2153	15,837	17,990	2	1	2060	12,944	15,004
34	1.09	22,695	1.09	1690	20,833	22,523	2	1.04	1611	19,857	21,468
35	1.08	21,239	1.04	2051	18,417	20,468	2	1	1976	17,744	19,720
36	1	13,157	1	917	12,240	13,157	2	1	917	12,240	13,157
37	1.36	14,671	1.36	1093	13,578	14,671	2	1	796	9980	10,776
38	1.25	16,130	1.05	1469	12,100	13,569	2	1.04	1453	11,971	13,424
39	1.42	14,203	1.2	1310	11,781	13,091	2	1	1092	8934	10,026
40	1.23	17,797	1.09	1619	14,136	15,755	2	1.04	1542	13,464	15,006
41	1.6	17,560	1.3	1700	14,276	15,976	2	1.14	1488	10,977	12,465
42	1.14	15,329	1	1352	12,121	13,473	2	1	1350	12,102	13,452
43	2.04	17,560	1.63	1450	15,219	16,669	3	1.24	1107	10,596	11,703
44	1.36	16,048	1.11	1244	14,451	15,695	3	1	1122	10,652	11,774
45	1.58	22,310	1.26	2344	16,321	18,665	3	1	1855	12,302	14,157
46	2.27	18,462	2.27	1437	17,002	18,439	3	1.49	1447	11,171	12,618

47	1.82	13,849	1.82	960	12,889	13,849	3	1.33	849	9426	10,275
48	1	8401	1	398	8003	8401	3	1	398	8003	8401
49	2.03	20,982	2.01	1390	19,589	20,979	3	1.15	1378	11,228	12,606
50	1	7263	1	681	6582	7263	3	1	681	6582	7263
51	1.65	19,755	1.47	1947	15,549	17,496	3	1.09	1582	11,544	13,126
52	1.83	19,440	1.81	1609	17,686	19,295	3	1.17	1509	11,372	12,881
53	2.45	21,139	2.42	1392	19,639	21,031	3	1.42	1498	11,492	12,990
54	2.18	11,991	2.17	908	11,083	11,991	3	1.44	710	7370	8080
55	1	12,936	1	1715	11,221	12,936	3	1	1715	11,221	12,936
56	1.27	10,387	1.26	900	9487	10,387	3	1	624	7552	8176
57	1.35	10,798	1.29	1411	9387	10,798	3	1.29	1411	9387	10,798
58	1	6631	1	819	5812	6631	3	1	819	5812	6631
59	2.04	19,392	1.73	1799	14,568	16,367	3	1.35	1508	11,431	12,939

Table 15.2 Role of assortment and regional differences

Paired sample t-*test: Test for assortment differences*				
Analysis	Mean Efficiency	n	t-value	p-value
Analysis 1	1.39	59	4.96	.000
Analysis 2	1.29	59		
Paired sample t-*test: Test for assortment differences*				
Analysis	Mean sales volume estimate	n	t-*value*	p-*value*
Analysis 1	17,854	59	5.29	.000
Analysis 2	16,939	59		
Paired sample t-*test: Test for regional differences*				
Analysis	Mean efficiency	n	t-*value*	p-*value*
Analysis 2	1.29	59	6.72	.000
Analysis 3	1.08	59		
Paired sample t-*test: Test for regional differences*				
Analysis	Mean sales volume estimate	n	t-*value*	p-*value*
Analysis 2	16,939	59	9.02	.000
Analysis 3	14,391	59		

the number increases to 30 in Analysis 3. Furthermore, the overall average inefficiency declines from 1.29 (Analysis 2) to 1.08 (Analysis 3). The decline in inefficiency is significant (Table 15.2: $t = 6.72$, $p < .001$). Additionally, the projected sales also decline ($t = 9.02$, $p < .001$).

15.5.3 Comparison of Regression-Based Approach to DEA-Based Approach

As mentioned earlier, regression-based sales estimates compare an individual store to the average performer, whereas the DEA approach compares an individual store to the best-performers. Thus, the regression-based approach is likely to provide more conservative sales estimates. To examine this issue, a regression model was run with the same variables as in the first DEA analysis—the independent variables were variable expenses, square footage, and inventory); whereas the dependent variable was sales. The resulting sales estimates from the regression analysis were compared to the Analysis 1 projections. As expected the DEA projections were larger (Table 15.3: $t = 22.42$, $p < .001$).

15.5.4 Slack Analysis

The slack analysis is provided in addition to the sales goals in Table 15.4. Recall that the sales goal of Product A and B are calculated as the projected sales for the retail outlets to be efficient. We calculated the slack as a percentage of current sales,

Table 15.3 Regression-based versus DEA-based estimates

Analysis	Mean sales volume estimates	n	t-*value*	p-*value*
Analysis 1 (DEA)	17,854	59	22.42	.000
Regression	13,454	59		

Paired sample *t*-test: Test for Differences Based on Methodology

i.e., percentage increase of current sales volume required to transform an inefficient outlet into an efficient outlet. The results on individual stores as well as the averages are presented in Table 15.4. The slack for both products A and B sales in Analysis 1 was 39.66 % and in Analysis 2 was 32.52 % ($t = 5.01$; $p < 0.001$). The slack percentage for Analysis 3 for both products A and B was 9.26 %, significantly different from Analysis 1 and Analysis 2 ($t = 8.87$ and 7.58, respectively; $p < 0.001$). The slack percentage was significantly different in Analysis 2 and Analysis 3 for Product A (32.84 % and 8.71 %, respectively, $t = 7.44$; $p < 0.001$) and Product B (30.84 % and 16.59 %, respectively, $t = 6.59$; $p < 0.001$). Therefore, the results of comparing slacks from the three analyses were similar to the results of the efficiency analysis.

15.6 Implications and Avenues for Future Research

The implications for evaluating performance at a more disaggregate level using a "best practice" methodology such as DEA are in two major areas. First, the method illustrates a method of goal setting and assortment planning that is superior in many ways to more traditional methods. Second, the results of the process provide directions for enhancing the overall performance of the retail chain. These issues and avenues for future research are discussed next.

15.6.1 Goal Setting and Assortment Planning

The issues of evaluation, goal setting and assortment planning are becoming more important for retailers because of the increase in competition and changing retail paradigms. For example, research on store evaluation and retail productivity has been prolific as seen by the special issues of *Journal of Retailing* in 1984 and 1997 (e.g., Achabal et al. 1984; Ingene 1984), and *International Journal of Research in Marketing* in 1997. The traditional measures of efficiency or retail productivity (i.e., a ratio of a single output to a single input) have made the evaluation of retail outlets in different regions that carry different assortments very difficult. In this paper, we presented a method of obtaining efficiency measurements (using multiple inputs and outputs) that compare individual outlets with "best practice" outlets located in the same region.

Table 15.4 Slack analysis

	Analysis 1	Analysis 2			Analysis 3		
Stores	Percentage slack for Products A and B	Percentage slack for Products A and B	Percentage slack for Product A	Percentage slack for Product B	Percentage slack for Products A and B	Percentage slack for Product A	Percentage slack for Product B
1	9.35	8.57	8.57	8.59	3.54	3.02	9.90
2	39.19	29.66	29.66	29.69	0.00	0.00	0.00
3	48.48	44.46	44.46	44.44	0.00	0.00	0.00
4	15.26	15.02	15.02	15.02	0.00	0.00	0.00
5	45.97	45.68	45.69	45.67	11.40	10.83	18.08
6	80.51	79.63	79.63	79.65	43.12	42.00	57.54
7	19.70	15.11	15.11	15.11	6.80	6.80	6.81
8	5.06	0.00	0.00	0.00	0.00	0.00	0.00
9	15.77	0.00	0.00	0.00	0.00	0.00	0.00
10	14.77	6.42	6.42	6.44	0.00	0.00	0.00
11	21.18	16.93	16.93	16.91	6.77	6.77	6.76
12	0.00	0.00	0.00	0.00	0.00	0.00	0.00
13	35.38	14.10	15.21	4.88	0.00	0.00	0.00
14	87.43	55.23	55.23	55.23	0.00	0.00	0.00
15	21.08	16.48	16.48	16.49	1.19	1.19	1.18
16	45.68	45.00	45.00	45.03	19.50	19.50	19.48
17	0.00	0.00	0.00	0.00	0.00	0.00	0.00
18	19.02	12.12	12.12	12.11	2.01	2.01	1.98
19	13.71	12.30	12.30	12.30	4.90	4.90	4.92
20	81.25	47.00	55.44	2.32	0.00	0.00	0.00
21	0.00	0.00	0.00	0.00	0.00	0.00	0.00
22	42.06	41.75	41.75	41.73	4.99	4.99	5.02
23	0.00	0.00	0.00	0.00	0.00	0.00	0.00

24	71.32	56.41	56.41	56.42	0.00	0.00	0.00
25	60.25	59.61	59.61	59.59	20.56	20.56	20.55
26	28.23	28.01	28.01	28.03	7.06	7.06	7.07
27	43.88	43.52	43.52	43.49	17.53	17.53	17.52
28	0.00	0.00	0.00	0.00	0.00	0.00	0.00
29	54.10	40.93	41.25	38.60	20.36	20.36	20.35
30	54.76	0.00	0.00	0.00	0.00	0.00	0.00
31	23.87	22.77	22.76	22.79	4.47	4.47	4.45
32	14.20	9.66	9.65	9.68	0.00	0.00	0.00
33	25.10	19.90	22.35	4.51	0.00	0.00	0.00
34	9.45	8.62	8.62	8.61	3.54	3.54	3.53
35	7.70	3.79	3.79	3.80	0.00	0.00	0.00
36	0.00	0.00	0.00	0.00	0.00	0.00	0.00
37	36.15	36.15	36.05	37.31	0.00	0.00	0.00
38	25.16	5.28	5.28	5.30	4.16	4.16	4.16
39	41.66	30.57	31.87	19.96	0.00	0.00	0.00
40	22.89	8.79	8.79	8.80	3.62	3.62	3.63
41	59.93	45.50	47.65	29.67	13.52	13.53	13.50
42	13.95	0.16	0.16	0.15	0.00	0.00	0.00
43	103.67	93.33	96.83	62.92	35.73	37.04	24.38
44	36.30	33.30	35.66	10.87	0.00	0.00	0.00
45	57.59	31.84	32.67	26.36	0.00	0.00	0.00
46	126.86	126.58	126.57	126.66	55.05	48.87	128.23
47	82.39	82.39	82.07	86.77	35.32	33.15	65.18
48	0.00	0.00	0.00	0.00	0.00	0.00	0.00
49	102.67	102.64	101.45	120.99	21.76	15.47	119.08
50	0.00	0.00	0.00	0.00	0.00	0.00	0.00

(continued)

Table 15.4 (continued)

Stores	Analysis 1	Analysis 2			Analysis 3		
	Percentage slack for Products A and B	Percentage slack for Products A and B	Percentage slack for Product A	Percentage slack for Product B	Percentage slack for Products A and B	Percentage slack for Product A	Percentage slack for Product B
51	65.42	46.51	46.51	46.50	9.91	8.77	19.04
52	82.60	81.24	81.25	81.19	20.99	16.54	69.93
53	144.52	143.27	142.46	155.41	50.26	41.88	174.86
54	117.82	117.82	117.02	128.14	46.78	44.31	78.39
55	0.00	0.00	0.00	0.00	0.00	0.00	0.00
56	27.04	27.04	25.62	44.23	0.00	0.00	0.00
57	34.98	34.98	35.93	28.98	34.98	35.93	28.98
58	0.00	0.00	0.00	0.00	0.00	0.00	0.00
59	104.41	72.52	72.52	72.48	36.39	35.37	44.58
Average	39.66	32.52	32.84	30.84	9.26	8.71	16.59

Slack Percentage = 100 $*$ (Target Sales−Current sales)/(Current Sales)

Retail chains recognize that regional idiosyncrasies should be included in the store evaluation process. Yet, traditional evaluation methods do not allow for these differences. Retail firms typically classify their stores into A, B, and C level stores based on sales of the store and the markets in which they operate. This ABC typology is carried through to the evaluation of categories. The DEA methodology used in this paper provides a more accurate assessment of the productivity of stores than the ABC approach because it controls for regional and assortment differences. For instance, the accumulated goodwill and familiarity of Nordstrom's in the Seattle area should lead to higher levels of performance and productivity than the new store in Denver, because Nordstrom's was founded in and has had several stores in Seattle for many decades. Thus, using traditional analyses, the performance of the A store in Denver would be unfairly compared directly with an A store in Seattle. In this study, store 20 in region 2 was rated as being extremely inefficient (score of 1.81) in the aggregate analysis, but was considered to be efficient when regional and assortment differences were taken into consideration. Thus, using a traditional aggregated analysis, store 20 and its manager could be unfairly penalized.

A key diagnostic benefit of the DEA methodology is the slack analysis. In particular, in this study we concentrate on output-oriented slack analysis. As mentioned earlier, the slack analysis (provided in Table 15.4) would allow the chain to understand what percentage increase in sales volume is required to transform an inefficient retail outlet to an efficient retail outlet. For example, in Analysis 3, store 6 (inefficiency rating of 1.42) would need to increase its target sales by 42 % for product A and 57.4 % for product B to transform it from an inefficient retail outlet to an efficient retail outlet.

Alternatively, using an input-oriented slack analysis, the retail chain could try and assess what percentage reduction in inputs would help transform an inefficient retail outlet to an efficient outlet. An input-oriented DEA analysis suggests that store 15 would need to reduce operating expenses by 13.68 %, square footage by 22.06 %, and inventory by 17.73 % to become efficient. Note, however, that a reduction in operating expenses and inventory are more controllable in the short term compared with square footage.

Another strength of the DEA methodology is that it can calculate a fair and realistic sales target for specific merchandise categories based on the region in which a particular store operates. For example, retailers know that it would not be fair to compare the sales of truck tires of a dealer in Wyoming with one in the greater New York City area. Yet, using traditional methodologies, retailers have no accurate basis of setting sales goals between regions.

DEA allows retailers to predict target sales for any unit of analysis from the SKU to the store level based on the "best practice" SKUs, categories, or stores in similar locations. Although regions were used as the basis of analysis in this application, other criteria, such as trade areas, or stores whose customers are psychographically similar could also be used.

DEA is very useful for determining merchandise budgets fairly. Because it predicts what sales should be for a particular category in a given store using the slack analysis, it provides accurate information to buyers and store managers about how particular categories are performing compared with what they should be doing. For instance, a traditional analysis would base the sales targets for two K-Mart swimwear departments in St. Petersburg and Clearwater Florida on past sales. Instead of using a backward measure like past sales, the DEA analysis looks forward into what sales should be. Suppose the DEA analysis identifies the Clearwater store as efficient, and the St. Petersburg store as inefficient. Because they are "comparable" stores, the DEA analysis would suggest that the St. Petersburg store sales predictions should be based on the efficient Clearwater store information.

Another problem with traditional methods that forecast future sales using a percentage increase of last year's sales is that it penalizes the better performers while giving below average stores an advantage. Consider again the two stores in St. Petersburg and Clearwater. St. Petersburg has sales this year of $50 million, whereas Clearwater has $80 million. Traditional planning may suggest a 10 % increase for both stores, forecasting sales of $55 million and $88 million, respectively. A DEA analysis may show that St. Petersburg is operating at 50 % efficiency and needs to improve to 80 % efficiency, setting a goal of $80 million. On the other hand, Clearwater may have an efficiency of 100 %, requiring a modest increase to $82 million. The DEA analysis, therefore, rewards the Clearwater store for doing well in the past by giving it a moderate increase, whereas raising the hurdle for the St. Petersburg store.

15.6.2 Directions for Enhancing the Overall Performance of the Retail Chain

Competition in retailing is increasingly being regarded as an information contest. Retailers, such as Wal-Mart, are admired more for their prowess in the information arena than for their ability to pick merchandise or locate stores. In fact, information has become so critical that Wal-Mart accused and sued Amazon.com for stealing their IT talent. Amazon.com and other virtual retailers are being viewed by Wall Street with intense optimism. Price/earnings ratios on these firms are astronomical.

Retailers have advanced information systems for everything from procuring merchandise to locating stores. They have invested millions in sophisticated data warehouses. For example, Wal-Mart has developed a 24-tetrabyte data warehouse. Sears and Kmart have 14- and 8-tetrabyte warehouses, respectively (Zimmerman 1998). The degree to which retailers are able to harness this information is expected to determine their success in the future.

The proposed methodology can be used to enhance the retail performance by examining and propagating 'best practice' skills. Once the DEA methodology identifies 'best practice' stores, top management should examine two aspects of their behaviors. First, they should determine what the more successful store managers and buyers "DO" when they face everyday situations. At a deeper level, they should determine how the more successful store managers and buyers "THINK" about stores and customers. The first exercise develops behavioral guidelines for the managers and buyers, whereas the second exercise develops cognitive or thinking guidelines about customers and stores.

Behavioral guidelines are broad lessons that store managers and buyers have learned from good and bad situations (i.e., what to do and what not to do). Previous research and experience have shown that the best way to learn behavioral guidelines is through story-telling, i.e., case studies of successful and unsuccessful situations (Klein 1998). The best retailers, e.g., Home Depot, JCPenney, and Wal-Mart, are known to disseminate this type of information during weekly meetings, closed circuit satellite broadcasts, and in-house publications. Some, e.g., Nordstrom, encourage innovative and heroic service efforts by having employees tell about their activities and giving awards for the best ones.

Cognitive guidelines provide inputs into how store managers should think about their stores and customers. Research suggests that in a store identified as a 'best practice' store, managers and their employees will classify and interact with their customers differently than those managers and employees in mediocre performing stores (c.f., Klein 1998). The way these managers and employees think is referred to as their knowledge structures. To enhance performance, organizations have found it useful to disseminate "knowledge structure" information from best practice managers and key employees (c.f., Klein 1998). In short, behavioral guidelines will lead store managers to understand that "behaviors 'x' and 'y' work with our customers", and cognitive guidelines will refine that understanding to determine what type of behaviors to use with different types of customers.

15.6.3 Avenues for Additional Research

The results of the paper suggest areas for additional research. First, the selection of standardized or similar inputs and outputs across stores are proposed. This would make it easier to compare chains. Although our selection of inputs and outputs were based on management input and past research, future studies should explore different input and output measures. Similarly, it is important to determine what factors are associated with 'best practice' stores, e.g., characteristics of target market, geographical similarities. Once known, managers can attempt to clone their best performers through organizational learning-based research. Finally, we chose to examine the effects of region and assortment on the merchandise planning process. Other factors, such as type of location or demographic/psychographic makeup of the trade area, may be equally important in other research settings. In the future, researchers should attempt to determine which are the most important factors to consider for various retail formats.

References

Achabal DD, Heineke JM, McIntyre SH (1984) Issues and perspectives on retail productivity. J Retail 60:107–127

Banker RD, Morley RC (1986a) Efficiency analysis for exogenously fixed inputs and outputs. Oper Res 4:513–521

Banker RD, Morley RC (1986b) The use of categorical variables in data envelopment analysis. Manage Sci 32:1613–1627

Bessen J (1993) Riding the marketing information wave. Harv Busi Rev. September–October:150–163

Blattberg RC, Deighton J (1991) Interactive marketing: exploiting the age of addressability. Sloan Manage Rev 33:5–17

Boles J, Donthu N, Lohtia R (1995) Salesperson evaluation using relative performance efficiency: the application of data envelopment analysis. J Pers Sell Sales Manage 15:31–49

Bucklin LP (1978) Productivity in marketing. American Marketing Association, Chicago

Chebat J-C, Filiatrault P, Katz A, Tal SM (1994) Strategic auditing of human and financial resource allocation in marketing: an empirical study using data envelopment analysis. J Bus Res 31:197–208

Donthu N, Yoo B (1998) Retail productivity assessment using data envelopment analysis. J Retail 74:74–92

Gilmore J, Joseph Pine B (1997) The four faces of mass customization. Harv Bus Rev 75:91–101

Ingene CA (1982) Labor productivity in retailing. J Mark 46:75–90

Ingene CA (1984) Productivity and functional shifting in spatial retailing: private and social perspectives. J Retail 60:15–36

Kahn B (1991) Consumer variety-seeking among goods and services. J Retail Cons Serv 2 (3):139–148

Kahn B, Lehmann DR (1991) Modeling choice among assortments. J Retail 67(3):274–299

Kahn B, McAlister L (1997) Grocery revolution. Addison Wesley, Boston

Kamakura WA, Ratchford BT, Agrawal J (1988) Measuring market efficiency and welfare loss. J Cons Res 15:289–302

Kamakura WA, Lenartowicz T, Ratchford BT (1996) Productivity assessment of multiple retail outlets. J Retail 74:333–352

Klein G (1998) Sources of power. MIT, Cambridge

Lusch RF, Jaworski BJ (1991) Management controls, role stress and retail store manager performance. J Retail 67:397–419

Lusch RF, Serpkenci RR (1990) Personal differences, job tension, job outcomes, and store performance: a study of retail store managers. J Mark 54:85–101

Mahajan J (1991) A data envelopment analytic model for assessing the relative efficiency of the selling function. Eur J Oper Res 53:189–205

Murthi BPS, Srinivasan K, Kalyanaram G (1996) Controlling for observed and unobserved managerial skills in determining first-mover market share advantages. J Mark Res 23:329–336

Parsons LJ (1990) Assessing salesforce performance with data envelopment analysis. Paper presented at TIMS marketing science conference. University of Illinois, Urbana

Ratchford BT, Brown JR (1985) Study of productivity changes in food retailing. Mark Sci 4:292–311

Ratchford BT, Stoops GT (1988) A model and measurement approach for studying retail productivity. J Retail 64:241–263

Solomon B (1994). Mass chains turning stocks store by store. Women's Wear Daily (June, 8): 1, 17

Yadav M, Monroe KB (1993) How buyers perceive savings in a bundle price: an examination of a Bundle's transactional value. J Mark Res 30:350–358

Zimmerman KA (1998) Kmart to quadruple size of data warehouse. Supermarket News 48(28):19

Chapter 16
Evaluation of Subsidiary Marketing Performance: Combining Process and Outcome Performance Metrics

Dhruv Grewal, Gopalkrishnan R. Iyer, Wagner A. Kamakura, Anuj Mehrotra, and Arun Sharma

Abstract Issues in evaluating marketing performance and devising appropriate metrics for measurement have taken center stage in marketing thought and practice in recent years. We propose an empirical model that enables a multinational enterprise (MNE) to assess the marketing performance of its subsidiaries, taking into explicit consideration the fact that tactical actions by subsidiaries contribute to the creation of assets that can be harnessed for marketing outcomes. Thus, our model captures the asset creation abilities of marketing expenditures and also takes in to account the environmental differences of the context in which each MNE subsidiary operates. We evaluate comparative, overall, and process-level (creation of market assets and market yield) marketing performance in the context of multi-

This chapter is reproduced from, Grewal, Dhruv, Gopal Iyer, Wagner Kamakura, Anuj Mehrotra and Arun Sharma (2009), "Evaluation of subsidiary marketing performance: combining process and outcome performance metrics," *Journal of the Academy of Marketing Science,* 37 (2), 117–129, with permission from The Academy of Marketing Science.

D. Grewal (✉)
Division of Marketing, Babson College, Babson Park, MA 02468, USA
e-mail: dgrewal@babson.edu

G.R. Iyer
Center for Services Marketing, Florida Atlantic University, 777 Glades Road, Boca Raton, FL 33431, USA
e-mail: giyer@fau.edu

W.A. Kamakura
Rice University, Houston, TX, USA

A. Mehrotra
Management Science, University of Miami, Miami, FL, USA

A. Sharma
University of Miami, Miami, FL, USA

J. Zhu (ed.), *Data Envelopment Analysis*, International Series in Operations Research & Management Science 238, DOI 10.1007/978-1-4899-7684-0_16

country operations. This simultaneous examination of marketing process and marketing outcome performance enables a global corporation to gain strategic, operational, and diagnostic insights into the performance of its subsidiaries. Our approach is empirically illustrated with an evaluation of the marketing performance of subsidiaries of a large global corporation.

Keywords Multinational performance evaluations • Marketing metrics • Outcome measures • Performance measures • Standardization

There is now more pressure on marketing scholars and practitioners to demonstrate that the marketing function contributes to shareholder value for the firm (Doyle 2000; Rust et al. 2004). The importance of justifying marketing investments and the metrics necessary to measure marketing performance thus have taken center stage. However, though extant conceptual models link marketing expenditures and tactical actions to the creation of marketing assets—which can be harnessed over the long-term (Rust et al. 2004; Srivastava et al. 1998, 1999)—few empirical studies relate marketing expenditures to marketing performance through the creation of market-based assets.

At the same time, the justification of marketing expenditures and the assessment of marketing performance is particularly complex for multinational enterprises (MNE). Although MNE performance assessment is clouded by various economic and accounting exposure risks, such as translation and transaction risks (Shapiro 2006), such firms must separate the unique contributions of the marketing functions conducted by their subsidiaries. Country-level operations augment the complexities of both the measurement process and the evaluation of global performance and therefore require unique adaptations to the design of marketing performance evaluation systems (Hewett and Bearden 2001). Further complicating these assessments, little consensus exists about how to measure the long-term effects of marketing activities (Dekimpe and Hanssens 1995). Measuring and reporting marketing performance to external stakeholders certainly is important (e.g., Doyle 2000), but global MNEs must also consider measures of marketing performance within and across their various subsidiaries.

Financial performance assessments of the overall corporation provided to shareholders are facilitated by legal and tax guidelines, but effective evaluations of marketing performance across the global organization, even for internal purposes, encounter several challenges. For example, the marketing performance of a subsidiary at any given time consists of the cumulative impact of various marketing processes and activities. Marketing measures that reflect only simple output–input ratios do not capture this complexity. Moreover, environmental factors such as competitiveness and regulatory constraints affect country-level operations, so these factors must be taken into account when comparing subsidiaries that operate in different national environments (Carpano et al. 1994; Porter 1986).

We examine country-level marketing performance within a framework of marketing processes and outcomes in the context of a global firm's country-level operations. Our framework explicitly considers the creation of market assets as an intermediate outcome between marketing expenditures and marketing performance and thus captures the impact of such long-term assets on marketing performance. As noted by Rust et al. (2004: p. 77), "marketing actions both create and leverage market-based assets." Thus, we posit that marketing performance of global subsidiaries relies on dual marketing processes within unique environments: market asset creation and market yield processes. Market asset creation refers to all activities undertaken by an organization to attract and retain customers, develop markets through advertising and distribution programs, and create and sustain brands. Market yield processes refer to the deployment of market-based assets. Considering these two processes, we propose a strategic classification matrix that provides a better understanding of the performance and operations of subsidiaries, given the environmental variations in which their marketing strategies get executed. We illustrate the model using data from the country-level marketing operations of a large MNE.

16.1 Marketing Performance Assessment in the Global Context

Empirical assessments of subsidiary performance within the international business and marketing literatures are relatively sparse. The neglect of the individual firm, as noted by Craig and Douglas (2000), may occur because researchers simply assume that leveraging a domestic positioning will help the firm succeed in international markets as well. Or it could be the lack of access to data on subsidiary operations.

Prior literature addresses the strategy–performance link in international business according to four perspectives: dynamic capability, standardization, configuration–coordination, and integration–responsive (Luo 2002; Zou and Cavusgil 2002). Each perspective yields insights into assessments of global marketing performance and contributes to the development of market creation and market yield processes as bases for evaluating subsidiary performance. The dynamic capability perspective calls for building and leveraging capabilities across the MNE network of subsidiaries (Luo 2002), so performance assessment involves creation-oriented and yield/exploitation processes. The standardization perspective entreats MNEs to seek scale economies by standardizing their marketing activities across subsidiaries and adapting marketing strategy to relevant environmental differences (Syzmanski et al. 1993). Integration–responsive (Zou and Cavusgil 2002) and configuration–coordination (Bartlett and Ghoshal 1989; Craig and Douglas 2000) perspectives instead suggest the need to leverage location-specific advantages and take explicit account of firm- and country-specific advantages enjoyed by each subsidiary while coordinating activities across subsidiaries to gain relevant synergies and control location-specific advantages and disadvantages.

Literature describing these four perspectives indicates several key conclusions. First, broadened measures of marketing performance should include both tangible and intangible factors (e.g., Bharadwaj et al. 1993; Srivastava et al. 1998). Second, marketing performance assessment must not only compare the relative performance of each subsidiary but also focus on processes that lead to superior performance (Birkinshaw and Morrison 1995; Tallman 1991; Zhao and Luo 2002), because marketing performance is not simply a direct relationship between inputs and outputs but involves developing enduring tangible (e.g., branded product lines) and intangible (e.g., brand awareness) market-based assets that can contribute to superior performance (Srivastava et al. 1998).

Third, traditional measures of performance should be reexamined to create better measurement models. The dynamic capability perspective, which addresses temporally and contextually contingent market-based assets, recommends developing marketing metrics within the context of the creation and deployment of existing market-based assets. Thus, marketing performance assessments must deal with how these assets prompt financial outcomes such as sales, profits, and cash flows (Srivastava et al. 1998). Marketing expenditures create market-based assets that can maximize the revenue and other outputs of long-term marketing investments. Keh et al. (2006) show in the context of services that raw marketing inputs may relate to final revenue outcomes through intermediate processes.

Fourth, in addition to a focus on assuring shareholders of marketing performance, measurement approaches should provide a diagnostic tool for the company. Performance assessment is a critical component of the marketing control process, so assessments should identify areas that need improvement, expenses to curb, and investments to make as well as provide a fair mechanism for evaluating divisional, segment, or, in the case of a MNE, subsidiary (Kim and Mauborgne 1993; Taggart 1997) performance.

16.1.1 Subsidiary Marketing Performance Assessment: A Model

As Zou and Cavusgil (2002) note, each perspective in international business literature offers a partial explanation of how to enhance marketing performance. An integrative approach should include not only direct impacts on performance but also processes that contribute to intermediate and final outcomes. Borrowing insights from the dynamic capability perspective, we take into account the processes of marketing asset creation and deployment and thus emphasize explicit contributions to marketing performance. With the standardization perspective, we can examine the same or similar variables across various subsidiaries; the integration–responsiveness and configuration–coordination perspectives reiterate

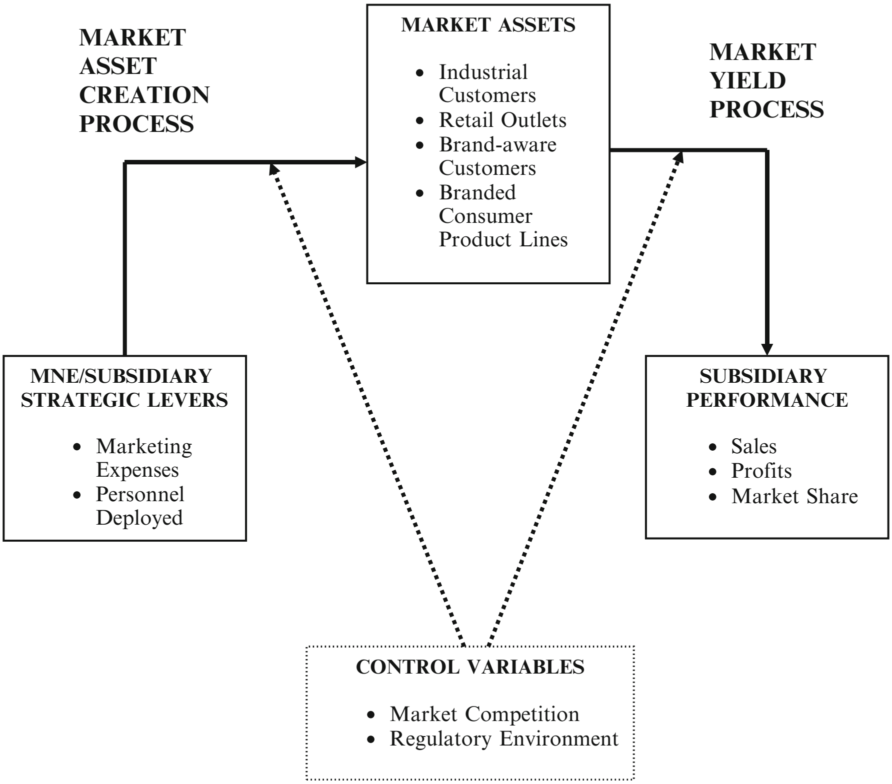


Fig. 16.1 Two-stage process for subsidiary marketing performance assessment (Note: **Market asset creation efficiency** relies on two inputs (marketing personnel and marketing expenditures) and three outputs (number of brand-aware customers, number of retail outlets, and number of industrial customers). **Market yield efficiency** uses three inputs (number of brand-aware customers, number of retail outlets, and number of industrial customers) and three outputs (sales, market share, and profits))

the need for appropriate environmental controls that may affect subsidiary operations and contribute to differential performance by subsidiaries of the same MNE.

To understand overall marketing performance, marketing operations could be considered as two natural sub-processes processes (market creation and market yield) for measuring the individual performance of each subsidiary (Fig. 16.1). Because a country's operations are contextually and temporally contingent, the MNE must examine its overall performance in the context of the process-level activities of creating and deploying market-based assets and seek to coordinate these activities globally to achieve synergy with subsidiaries (Zou and Cavusgil 2002). We elaborate on these processes with specific reference to global marketing operations.

16.1.2 Market Asset Creation Process

Although market assets are created through a variety of strategic and tactical marketing actions, the focus is on marketing expenditures and their impact on market asset creation (Rust et al. 2004). Strategic and tactical marketing activities may include personal selling, advertising, and distribution (Anderson 1995), designed to create market-based assets such as distribution networks, brand awareness, brand preference, and a customer base. Marketing expenditures on strategic and tactical marketing activities not only affect immediate sales, such as those from a short-term promotion campaign, but also develop market assets that are more long term in their nature and impacts (Kamakura et al. 1991; Reinartz and Kumar 2003).

These marketing expenditure effects on the creation of market-based assets can be observed within each national context of the MNE's operations. Within each country, the subsidiary expends financial and personnel resources to create a marketing network and other relational market-based assets. Successful market creation efforts enable an extensive marketing network, the fruits of which can be measured by, among other things, the brand presence in country markets and the development of key relational accounts (Yip and Madsen 1996).

16.1.3 Market Yield Processes

Market yield activities aim predominantly to convert market assets into increased sales and profits. Firms differ in terms of how effectively they can harness the assets at their disposal. In the case of an MNE, different subsidiaries may have different assets, perhaps due to the environment in which they operate or their efforts to translate marketing expenses into long-term, market-based investments.

A MNE's global operations consist of subsidiaries operating in different environments with possibly different priorities to capture sales and revenue from existing market assets. Newer subsidiaries or those in markets with tremendous untapped potential may focus on creating market assets; those with a well-developed marketing network or that operate in mature markets might focus on capturing value from assets already developed. When a sufficient market-based asset network exists, the subsidiary's attention shifts to asset deployment or utilization, which involves optimizing all sources of rent, including additional shares of customers' wallet, improved profit margins from marketing activities, and a growing share of the overall market. As the market approaches saturation, the marginal cost of acquiring new customers increases, so the subsidiary's attention shifts to market yield activities, because market asset creation activities require tremendous efforts and expenditures (e.g., extending the brand to new markets).

Market yield processes also involve translating the subsidiary's asset-building efforts into asset-deployment efforts. Yet the success of a firm's marketing strategy and its resultant development of a marketing network and loyal customer base do not automatically confer enhanced sales, profits, or market share. Instead, the firm operates in a dynamically changing environment, and the market knowledge and lessons learned aid its market yield.

16.1.4 Strategic Classification

Traditionally, marketing research examines either the overall performance of the firm's marketing function or an individual element. For example, research on portfolio models (Wind et al. 1983) and market orientation (Kohli and Jaworski 1990) discusses the role of marketing within the functions of the firm. Similarly, research that models the pattern of responses to direct marketing campaigns (Basu et al. 1995) examines processes within the marketing function.

Recent research suggests that the links among marketing expenditures, the creation of market assets, and market yield may be quite complex (Rust et al. 2004). Marketing expenditures may not lead to the strategic development of market assets, or market assets may not be adequately harnessed to yield the desired marketing performance objectives, such as sales and profitability (Sheth and Sisodia 2002). Both the effectiveness and the efficiency of marketing activities may be called into question when evaluating these links (Keh et al. 2006; Sheth and Sisodia 2002). For example, Dowling and Uncles (1997) challenge the effectiveness of marketing expenditures, especially on customer reward programs, for developing a loyal customer base. Market creation efforts may not produce the desired results if the strategies are unsuccessful or inefficient because of, say, poor implementation or environmental factors. Reinartz and Kumar (2000) show that loyal customers are not always the most profitable, especially if the costs to serve them are not less than those for new customers. That is, market yield efforts might not contribute to the desired performance objectives, such as profitability, because of misdirected marketing efforts or environmental factors (Dowling and Uncles 1997).

In global operations, the effectiveness and efficiency of market asset creation and market yield processes result from not only strategic directions or tactical implementation but also the specific national environmental context of each subsidiary. A firm may expend considerable resources on marketing activities but not be able to create a strong brand in a specific country, especially if customers have low switching costs and/or regulations curtail the marketing mix activities in which the firm can engage. In emerging markets, for example, the absence of long-term contracts facilitates customer switching among mobile telephone service firms. In Scandinavia, restrictions on advertising affect the range of options for marketers and curtail market asset creation.

Any assessment of the marketing performance of subsidiaries, especially market assets and market yield efforts, must partial out the effects of environmental differences, including competition, regulation, and the country infrastructure necessary for marketing. Thus, instead of a simple relation between inputs and outputs, performance evaluation must consider the specific objectives of the subsidiary as

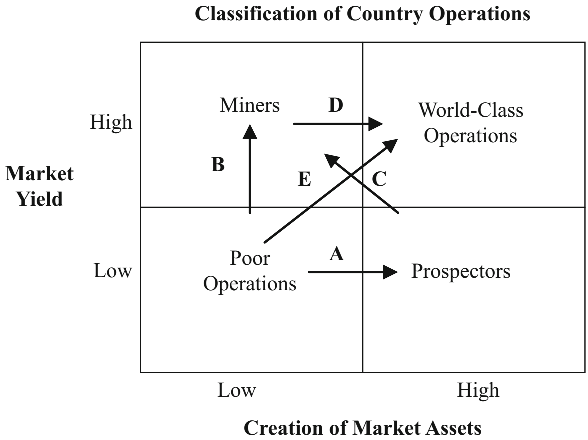


Fig. 16.2 Country-level operations matrix. (**Note:** Paths to superior performance: *A* primary focus on creation of market assets, *B* primary focus on market yield processes, *C* focus on market yield from existing accounts, *D* focus on creation of market assets to supplement market yield processes, *E* simultaneous focus on market creation and market yield processes)

well as the unique environment in which it operates. For example, MNE subsidiaries in all countries may not focus on the same marketing objectives. Some may be sales subsidiaries and perform an outpost function in which they coordinate independent distributors and service contractors. Others may be engaged in domestic account building activities.

Our examination of performance in terms of market assets and market yield enables the classification of country operations into four cells (Fig. 16.2). Starting from the upper right-hand quadrant, the first cell is "world-class operations"; country operations manage both the creation of market assets and market yield very effectively. They represent the best asset creation and deployment practices, and learning organizations should use them as standards to be emulated.

Moving counterclockwise, the second cell, "miners," represents country operations that manage market yield (deploy assets) very effectively but are unable to create market-based assets efficiently. Firms engaged in the short-term maximization of sales, especially through strategies such as price promotions, may fall into this cell. Their activities and expenditures produce results, but in the long term, their lack of attention to enduring market-based assets compromises their strategic position. Such subsidiaries are especially vulnerable to environmental threats, including changes in the competitive structure, and a well-positioned competitor with a clear focus on market-based assets can hinder their ability to optimize sales even in the short term.

The third cell is labeled "prospectors," whose country-level operations display high asset creation performance but low market yield (asset deployment). Prospectors focus on customer acquisition indiscriminately but obtain low yields from each customer for several reasons. First, the economic conditions in the target market may be so unfavorable that customers buy smaller quantities of the product.

Second, an emphasis on customer acquisition may lead to marginal customers who are light users of the product and less likely to convert into loyal customers. Third, trade decisions, such as retailers' display decisions, could influence or negate high brand equity in that country (Buchanan et al. 1999).

Subsidiaries falling into the fourth cell, "poor operations," require extensive attention to bring about improvements in both their asset creation and market yield (asset deployment) processes. Poor operations may result from improper marketing processes or unfavorable environmental conditions (e.g., economic or political stress).

The paths A through E in Fig. 16.2 identify the ways in which subsidiaries falling short of world-class operations could improve their performance. Poor operations result from poor performance in both market creation and market yield activities, so subsidiaries should focus on either the creation of market assets through suitable expenditures to generate new accounts (Path A) or improving market yield from existing accounts (Path B)—or both (Path E). Prospectors need to focus on improving their market yield processes (Path C), and miners must supplement existing accounts with new ones (Path D) to become world-class.

We do not classify the entire MNE into performance assessment cells; it may have world-class operations compared with its competitors, but its subsidiaries still may fall into another cell. Coca-Cola's subsidiaries demonstrate world-class performance in Mexico (highest per capita consumption) but prospector status in Peru (Inka Cola is the top-selling brand). Dell demonstrates world-class performance in the United States but is a miner in India.

The classification provides a static view of subsidiary performance. If all subsidiaries were established at the same time, faced the same market and environmental conditions, and provided similar inputs, it would be a snapshot of the absolute performance of each. However, subsidiaries inevitably face different market and environmental conditions, and the MNE evolves by establishing new subsidiaries. If in the MNE evolution, all subsidiaries may be considered "mature" because they have been in operation for several years, the classification provides a good summary of relative performance. Market and environmental differences also can be controlled for in the analytical model.

Finally, the classification is merely a diagnostic tool. Subsidiaries could move from being prospectors to world-class operations or, with sufficient impetus, to miners. The snapshot of performance at one point of time thus offers subsidiaries recommendations for improving their performance so that they may move to a different performance quadrant.

16.1.5 Combining Overall Marketing Performance and Marketing Process Performance

If the overall performance variable adequately captures the processes that underlie the relationships between overall inputs and outputs, country-level operations could be classified into two cells: high market asset creation with high market yield and low market creation with low market yield. However, marketing operations across countries are much more complex. Subsidiaries that perform better than their counterparts but enhance their performance by improving market assets or market yield can be classified as either low market creation with high market yield or high market creation with low market yield.

16.2 Measuring Overall and Process-Level Performance

Many traditional methods examine the performance of subsidiaries; we concentrate on comparative efficiency, which acknowledges both inputs and outputs and takes into account the operating unit's productivity in transforming inputs into outputs relative to other units that operate in similar conditions. The simplest way to measure efficiency is through productivity ratios. However, these simple output/input ratios are not directly applicable to most marketing operations that involve multiple inputs and outputs, unless we apply arbitrary weights to combine them into a single productivity index (Kamakura et al. 1996).

To accommodate multiple inputs and outputs, we evaluate subsidiaries using data envelopment analysis (DEA), a linear programming approach that identifies a piecewise-linear Pareto frontier that defines the most efficient transformation of multiple inputs into multiple outputs and then measures the efficiency of the subsidiaries relative to this efficiency frontier. Subsidiaries located at the frontier are designated the "best practices" against which we compare all other units. We thus can evaluate the efficiency of the other units in relation to the best practices located in the closest facet of the Pareto frontier.

Marketing literature uses DEA widely to assess the efficiency of sales branches (Mahajan 1991), banks (Kamakura et al. 1996), managerial performance (Murthi et al. 1996), and retail outlets (Grewal et al. 1999). Donthu et al. (2005) use DEA to benchmark marketing productivity. Because this technique is not new, we provide only a basic discussion as it applies to our case. Suppose we are interested in measuring the efficiency of subsidiary o relative to all subsidiaries $k = 1, 2, \ldots, K$. Each subsidiary k uses inputs x_{ik} ($i = 1, 2, \ldots,$ I) and allocative inputs z_{lk} ($l = 1, 2, \ldots,$ L) to produce outputs y_{jk} ($j = 1, 2, \ldots,$ J). The only distinction between inputs x_{ik} and allocative inputs z_{lk} is that the former are under the direct control of the subsidiary, whereas the latter are not, because they either are managed at corporate HQ or represent uncontrollable environmental factors. Charnes et al. (1978) demonstrate that the technical efficiency T_o for the subsidiary in question can be measured by solving the following linear programming problem:

$\min T_o$ subject to

$$y_{jo} \leq \sum_k \lambda_k y_{jk} \quad \text{for all outputs } j; \tag{16.1}$$

$$T_o x_{io} \geq \sum_k \lambda_k x_{ik} \quad \text{for all inputs } i; \tag{16.2}$$

$$z_{lo} \geq \sum_k \lambda_k z_{lk} \quad \text{for all allocative inputs } l; \tag{16.3}$$

$$\lambda_k \geq 0 \quad \text{for all subsidiaries } k; \text{ and } \sum_k \lambda_k = 1 \tag{16.4}$$

The convex combination of all subsidiaries defined by $\sum_k \lambda_k = 1$ represents the best practice with which the focal subsidiary o is compared. If $\lambda_o = 1$, the focal subsidiary o is at the Pareto frontier and therefore efficient ($T_o = 1$). Otherwise, all subsidiaries k for which $\lambda_k > 0$ define a facet of the Pareto frontier against which the focal subsidiary is compared, and T_o measures the degree of relative inefficiency of the focal subsidiary. In other words, $T_o < 1$ indicates that a combination of subsidiaries (defined by $\lambda_k > 0$, $k = 1, 2, \ldots, K$) can produce the same level of outputs as the focal subsidiary using only a fraction (T_o) of its inputs. The solution of the linear programming not only measures the efficiency of the focal subsidiary but also identifies the other subsidiaries that form the benchmark.

This interpretation holds only when constraints 1–3 are binding (i.e., $y_{jo} = \sum_k \lambda_k y_{jk}$, $T_o x_{io} = \sum_k \lambda_k x_{ik}$, and $z_{lo} = \sum_k \lambda_k z_{lk}$) or, in other words, when the outputs produced and allocative inputs consumed by the focal subsidiary are equal to those from the benchmark and the inputs from the focal subsidiary, after the efficiency adjustment, are equal to those from the benchmark. If output constraint 1 is not binding (i.e., $y_{jo} < \sum_k \lambda_k y_{jk}$), the focal subsidiary not only would need to use a fraction (T_o) of its current inputs but also would have to produce more of the j-th output until it equaled the level produced by the benchmark. An equivalent conclusion exists when an input constraint is not binding.

The DEA approach yields managerially useful byproducts. In addition to developing relative efficiency ratings for subsidiaries, the efficiency models provide information about the extent to which each input consumed by an inefficient subsidiary should be reduced to maintain the same level of output and keep the input mix ratios (i.e., the basic technological recipe) approximately unchanged. However, the DEA methodology has limitations. First, as a deterministic linear programming approach, the model does not take into account any measurement error in the inputs or outputs and is susceptible to outliers. This limitation can be overcome by superimposing a measurement model on the frontier estimation model, with a stochastic DEA extension (Cooper et al. 1998) or frontier estimation model (Greene 2000). However, for our application, this superimposition would require either longitudinal data to estimate a stochastic frontier or a strong set of assumptions about the nature of the measurement error. Second, the dimensionality of the Pareto frontier increases with the number of inputs and outputs, making it easier for a subsidiary to be placed at the frontier. However, this problem is not unique to the DEA methodology but common to most measures of relative efficiency.

16.3 Empirical Test

We illustrate our proposed framework for subsidiary assessment on data collected from 18 subsidiaries of a global *Fortune* 50 firm that sells automotive products to industrial and retail customers, has annual worldwide sales that exceed $100 billion, and employs more than 85,000 people. The subsidiaries are located in a geographical domain with relatively common cultural characteristics. The names of the company and its subsidiaries are not reported to ensure confidentiality. Also, all subsidiaries are engaged in retail operations to industrial and retail customers. By comparing the performance efficiency of a set of subsidiaries that are homogenous in their mandates from the MNC headquarters, we can compare them along the same dimensions. If different subsidiaries have different mandates or roles and these assignments are known a priori, the designated roles can be incorporated easily as allocative inputs in Eq. 16.3. Alternatively, we could compare the efficiency measures to the assigned roles a posteriori, in which case we would expect that the centrally assigned roles for each subsidiary directly explain inefficiencies in market creation or development.

To determine overall efficiency, we use the total number of marketing personnel and marketing expenses as inputs and sales volume, market share, and profits as outputs (Kamakura et al. 1996). We measure marketing expenses in dollars spent on marketing activities. Marketing personnel deployment is obtained from a comprehensive human resources study of the firm, in which each employee filled out a form to detail the number of hours he or she spent on all activities. The number of marketing personnel equals the sum of the proportion of employees' time (in human years) used for marketing activities. Measures of sales volume, market share, and profits come from company records.

To determine the efficiency of market asset creation and market yield, we obtain data on the intermediate process outputs. Market asset creation—attracting new customers, retaining existing customers, and laying the foundation for developing future customers—entails three variables: number of industrial customers, number of retail outlets, and number of brand-aware customers (these tap both customer and brand assets). The company provided information for these variables, as well as the population of each country. Market asset creation efficiency relies on two inputs (marketing personnel and marketing expenditures) and three outputs (number of brand-aware customers, number of retail outlets, and number of industrial customers). We obtain the number of brand-aware customers by multiplying unaided brand recall by the population. Market yield efficiency uses three inputs (number of brand-aware customers, number of retail outlets, and number of industrial customers) and three outputs (sales, market share, and profits).

To assess the creation of market assets and overall efficiencies, we may need to control for certain country- or market-specific variables outside the scope of the firm (Pilling et al. 1999). As suggested by the firm's managers, we control for market competitiveness and level of market regulation and thereby control for market-level factors that cause subsidiaries to operate in unequally attractive markets. Those with less competition should be more attractive.

Market condition variables (e.g., market competitiveness) should affect both overall and market creation efficiency ratings. Market pressures (i.e., competition, government control) also should diminish overall and in terms of market creation activities compared with when there are no market pressures. Using a modified Delphi technique, we asked marketing managers at the regional HQ to provide independent ratings of market competitiveness and the level of market regulation on 100-point rating scales. The ratings were circulated, and the group met to arrive at a consensus. An evaluation of 0 on each variable denotes high levels of market competitiveness and market regulation, which reflects allocative inputs (i.e., low values represent less productive environments). The data appear in Table 16.1, disguised through hidden conversion factors.

16.4 Results

In Table 16.2, we summarize the efficiency measures obtained with DEA. Using the overall and process-level efficiency measures, we classify each subsidiary into the strategic matrices by (overall) performance, creation of market assets, and market yield. As Fig. 16.3 shows, subsidiaries at the efficiency frontier (efficiency = 1.0) are classified as high; all others are low.

One of the advantages of using control variables is that factors outside the control of country-level operations appear in the DEA analysis. We control for market competitiveness and market regulation when assessing both overall and market asset creation efficiency. In Table 16.2, efficiency scores without control variables (in parentheses) differ from those obtained when we account for these environmental factors. Several interesting observations result. According to the overall efficiency ratings, subsidiaries 1, 4, 5, 6, 7, 8, 9, 12, 13, 15, 17, and 18 are at the frontier. According to the decision matrix in Fig. 16.3, subsidiaries 1, 4, 5, 6, 7, 8, 9, 12, 13 and 17 are world-class operators because they are at the Pareto frontier in both processes. Subsidiary 18 is efficient in the creation of market assets but inefficient in market yield and therefore is a prospector. Subsidiary 15 is a miner because it is efficient in market yield but inefficient in creation of market assets. For obvious reasons, none of the efficient subsidiaries are classified into the poor operations cell.

Subsidiaries 2, 3, 10, 11, 14, and 16 are inefficient with regard to overall marketing operations. Among them, units 2, 3, and 11 are inefficient in both market yield and market asset creation (poor operations). Subsidiaries 14 and 16 efficiently create market assets but deal with market yield inefficiently (prospectors). Subsidiary 10 has inefficient market asset creation but efficient market yield processes (miner). No subsidiary in this group is a world-class operation.

The classification matrix is diagnostic in terms of the process-level strategies that subsidiaries should follow. We discuss some of these strategies in the context

Table 16.1 Inputs and outputs for 18 subsidiaries

Country	Personnel (in human years)	Expenses	Competitiveness (100 points)	Regulation (100 points)	Sales volume	Market share	Profits	Retail outlets (number)	Industrial customers (number)	Brand-aware customers (number)
1	10.8	9	75	15	7	53	1.8	33	90	89.1
2	4.5	4	75	15	1.3	15	.6	12	38	47
3	7.4	2.9	75	15	.6	20	.4	18	47	94.7
4	7.1	6.3	75	15	3.3	80	2.5	14	190	23.2
5	2.8	2.3	75	15	.6	26	2	9	19	15
6	151.6	48.3	15	75	29.9	16	14.1	243	1322	2524.7
7	145.7	45.1	15	15	51	20	2.5	364	929	4007
8	34.3	11.5	50	15	15.3	22	5.3	103	109	1864.8
9	46.9	11.1	50	75	10.5	38	17.4	61	572	3432.6
10	21.4	12.4	50	50	3.8	13	2.3	45	3	270
11	55.6	16.1	50	50	7.7	19	4.5	111	310	2436.2
12	7.6	3.7	50	75	3.2	36	1.3	11	49	340.2
13	23.4	6.2	50	15	3.7	20	.6	68	88	1139.7
14	25.2	9	50	75	7.8	16	.2	68	288	617.3
15	42.3	8.4	50	50	8.9	62	5.1	50	151	535.7
16	18.1	9.8	50	75	4.6	7	2.1	52	101	634
17	74.3	42.3	15	75	23.3	11	4.3	279	271	121.3
18	10.5	1.1	75	15	1.7	15	0	31	41	121.3

Table 16.2 Efficiency and super-efficiency measures

	Efficiency			Super-efficiency		
	Overall	Creation	Yield	Overall	Creation	Yield
1	1.00	1.00 (.95)	1.00	1.43	.95	1.79
2	.82	.89	.90	.82	.89	.90
3	.68	.87 (.85)	.50	.68	.85	.50
4	1.00	1.00	1.00	---	2.32	---
5	1.00	1.00	1.00	2.39	1.69	5.69
6	1.00	1.00	1.00	---	---	---
7	1.00	1.00	1.00	---	---	---
8	1.00	1.00	1.00	1.30	1.22	2.03
9	1.00	1.00	1.00	---	3.11	---
10	.80 (.39)	.76 (.62)	1.00	.39	.62	12.86
11	.45 (.40)	.86	.43 (.40)	.40	.86	.40
12	1.00	1.00 (.92)	1.00	1.14	.92	1.60
13	1.00 (.51)	1.00	1.00 (.38)	.51	1.08	.38
14	.74 (.73)	1.00	.70 (.57)	.73	1.18	.57
15	1.00	.74 (.52)	1.00	1.20	.52	1.85
16	.62 (.54)	1.00 (.95)	.56 (.52)	.54	.95	.52
17	1.00 (.80)	1.00	1.00	.80	1.47	13.31
18	1.00	1.00	.70 (.55)	2.63	3.41	.55

Notes: Values within parentheses are efficiency scores without control variables
For illustrative purposes, the super efficiencies were calculated for the models without control variables. We used the DeaFrontier software to calculate the super efficiencies (see Zhu 2003)
--- indicates that it was not feasible to calculate super efficiency (see Seiford and Zhu 1999)

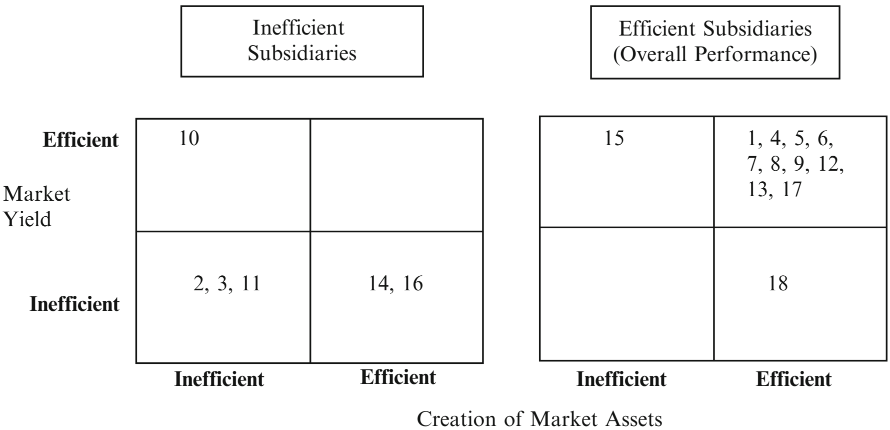


Fig. 16.3 Classification of country marketing operations (Note: Inefficient subsidiaries are those not on the efficiency frontier and efficient subsidiaries are those on the efficiency frontier)

of efficient subsidiaries, but the recommendations for inefficient subsidiaries are similar. Subsidiaries in the world-class operations cell should be studied and emulated by other firms. Subsidiary 18 is efficient in its overall marketing and creation of market assets but inefficient in its market yield, a mix that suggests it seeks marginal customers and instead should increase its customer selectivity. In contrast, subsidiary 15 is efficient overall and in market yield but not in creating market assets, which indicates that it should improve its ability to acquire new customers.

Recent research uses a super-efficiency DEA model to discriminate among efficient decision making units (DMUs) such as subsidiaries (Andersen and Petersen 1993; Banker et al. 1989). The super-efficiency results in Table 16.2 demonstrate that even among efficient subsidiaries, some are more efficient than others. However, several researchers question the use of super-efficiency model for ranking DMUs and instead advocate its use for other purposes, such as identifying influential observations (Banker and Chang 2006; Seiford and Zhu 1998).

Nevertheless, using only the efficiency measures, the DEA analysis provides data on a smaller subset of comparable units that a subsidiary should emulate. This diagnostic information appears in Tables 16.3 and 16.4 for subsidiaries that are inefficient in the creation of market assets and market yield, respectively. For example, subsidiary 15, classified as a miner, has an efficiency rate of .74 for market creation (Table 16.3). That is, a “virtual” subsidiary formed by a convex combination of units 9, 12, and 13, with weights of 15.4 %, 29.7 %, and 54.9 %, respectively, produces at least the same level of outputs as subsidiary 15 but uses only 74 % of its input levels. The results from Table 16.3 also indicate that a virtual subsidiary that defines the best practice could reduce its marketing personnel by 8.96 units (beyond the 74 % of the current level of inputs of subsidiary 15) and simultaneously increase its number of brand-aware customers by 720 units. In other words, to operate at the same performance level as the virtual benchmark, subsidiary 15 must maintain the same market creation outputs with 74 % of its current marketing effort, reduce marketing personnel by an additional 8.96 units, and increase brand-aware customers by 720 units. Comparing an inefficient unit with the relevant facet of the efficiency frontier thus provides clear directions for improving its performance in terms of the inputs to reduce and outputs to be produced at a higher rate. The analysis also indicates the efficient units that can engage in best practices for emulation. However, the DEA analysis cannot tell a manager how to achieve the necessary input reductions and output increases; this effort must be learned from an in-depth study of the best practices that form the virtual subsidiary for the focal unit.

Table 16.3 Inefficiencies in market creation

		Inefficient subsidiaries in market creation								
		2	3	10	11	13	14	15	16	18
Efficiency		0.89	0.87	0.76	0.86	1.00	1.00	0.74	1.00	1.00
Inputs	Personnel				1.25			8.96		
	Expenses	0.48		2.86						
Outputs	Retail outlets		0.66							
	Industrial customers			69.18						
	Brand awareness			304.20				720.00		
Allocative	Market competitiveness	0.37	1.03		3.06					
	Regulatory environment				18.00			7.94		
Frontier	Subsidiary 1	0.05								
	Subsidiary 4	0.08	0.11							
	Subsidiary 5	0.85	0.55	0.08						
	Subsidiary 7				0.09					
	Subsidiary 8	0.02			0.55					
	Subsidiary 9				0.28			0.15		
	Subsidiary 12			0.53				0.30		
	Subsidiary 13		0.04	0.34	0.08			0.55		
	Subsidiary 17			0.06						
	Subsidiary 18		0.30							

Table 16.4 Inefficiencies in market yield

		Inefficient subsidiaries in market yield								
		2	3	10	11	13	14	15	16	18
Efficiency		0.94	0.5	1	0.43	1	0.7	1	0.56	0.7
Inputs	Retail outlets									7.26
	Industrial customers		4.5				79.31			
	Brand awareness	22.62	32.35		515.5					
Outputs	Sales				0.33					
	Market Share	16.68	6		36.93		35.87		27.01	13.39
	Profit	1.42	1.6				3.89			2.09
Allocative	Market competitiveness									0.85
	Regulatory environment						18.98		10.68	
Frontier	Subsidiary 1	0.08					0.06		0.02	0.1
	Subsidiary 4	0.06								
	Subsidiary 5	0.85	1							0.87
	Subsidiary 7									
	Subsidiary 8								0.01	0.03
	Subsidiary 10				0.05				0.22	
	Subsidiary 12				0.07		0.28		0.61	
	Subsidiary 13				0.05					
	Subsidiary 15				0.84		0.61		0.13	
	Subsidiary 17						0.04		0.02	

16.5 Conclusion and Future Research Directions

16.5.1 Managerial Implications

This study presents a conceptual framework that can aid MNEs in evaluating subsidiary marketing performance in a multi-country context. The approach we present represents a marked departure from extant literature in international business and marketing. Whereas prior studies make comparisons across a broad spectrum of MNEs, our approach applies to the subsidiary level of an individual firm. It has more practical insights, because any MNE can use it to assess the marketing performance of its subsidiaries. It yields better comparative performance assessments, because the performance evaluation benchmarks refer to the unique context and operations of the MNE firm rather than other firms that operate in different industries, countries, or environmental contexts and with different constraints. Furthermore, our explicit consideration of process and environmental factors considers differences in the market environments across subsidiaries. The implications thus apply across a gamut of marketing decisions, including planning, analysis, implementation, and control of marketing activities.

Whereas overall performance measures actual, not potential, performance, the four-way classification we use can help identify subsidiaries with latent performance potential. By focusing on marketing performance processes, this approach supersedes conventional measures (e.g., overall marketing efficiency measurements, portfolio approaches) because it helps identify the process that needs improvement at each subsidiary level. In addition to providing assistance in resource allocation decisions, our approach identifies the exact process that needs specific resources. For example, rather than simply recommend increased budgets or resources, it helps MNE headquarters (HQ) identify specific activities to which to deploy resources to enhance performance. It also helps HQ identify extraneous sources of performance differences, such as distinct environmental variables. Performance assessments in overall marketing efficiency studies and portfolio models mistakenly assume similarity across units and attribute performance differences solely to factors beyond the control of the units themselves. Finally, our classification and performance assessment model is more actionable because it can identify specific strategies for performance improvements, whereas marketing efficiency studies only identify performance differences, and portfolio models help only with broad resource allocation decisions, such as decreasing investments, or irreversible decisions, such as exit.

To tackle the crucial problems of resource allocation and performance rewards, we recommend comparing subsidiary operations along the process priorities of creating market assets and obtaining market yield. In addition to its strategic implications, the framework reduces the adverse effects of relying on a single measure of performance and provides a greater degree of strategic control for the subsidiary, which leads to greater consensus in decision making and evaluation and, possibly, greater subsidiary compliance (Kim and Mauborgne 1993). Although we present our framework with cross-sectional data, managers also can use it to track relative changes in year-to-year performance and thus gain a better understanding of relative changes in market creation, market yield, and performance.

Therefore, MNEs that use the proposed framework and method will not only allocate resources across their various subsidiaries better but also leverage their deployed resources more effectively by providing specific strategic guidelines to subsidiaries. The HQ can mandate that subsidiaries focus on market creation or market yield activities to improve their marketing performance. Budgeting and resource allocation decisions become less ambiguous, following from a rational framework. Moreover, the use of a transparent analytical technique by HQ removes any potential political or other influence charges from budgeting decisions. Such transparency also has advantages in terms of the compensation of foreign managers, in that pay-for-performance compensation schemes can be executed and accepted across the MNE.

The results of our analysis could reveal the various marketing processes of the most efficient subsidiaries. With closer inspection, the MNE can identify specific strategic and operational processes that might be learned across the global corporation. These results and implications are similar to the standardization–adaptation decision and suggest that firms standardize in some countries to enhance efficiency and adapt in others to respond effectively to environmental variations (Szymanski et al. 1993). The results also enable a ready identification of subsidiaries, unique to the context of the firm, that could be designated "centers of excellence" on the basis of their strategic choices, implementation, and performance (e.g., Frost et al. 2002).

16.5.2 Limitations and Avenues for Further Research

Because our model can accommodate a variety of input and output variables, one of its limitations involves the constraints on variable selection and data collection. Whereas the model's flexibility in incorporating several variables is an asset, that asset can become a limitation because managerial discretion and even consensual decision making between HQ and various subsidiaries may be needed to identify the relevant variables.

At another level, there are inherent difficulties in identifying various contextual or control variables. For example, the level of competitive activity within a national environment may affect subsidiary performance, but measuring such competitive activity may be problematic. Objective data, such as the number of competitors or competitor sizes, market shares, and expenditures, may only be a start and need to be refined by qualitative judgments about the levels and extent of competitive activity. In our study, we overcome this limitation only through a separate survey that sought a consensual measure of competition in each national environment.

Further research might examine the extent to which country market experience and the stage of the product lifecycle in a specific country affect a subsidiary's strategy choice. Although all the subsidiaries in our study were established several decades ago, other MNE contexts might provide cases in which some subsidiaries are new and others quite mature. The age of the subsidiary and the product lifecycle stage should be modeled more explicitly in further research.

Additional research should take into account the relative differences in competitive positioning across various subsidiary operations. For example, subsidiaries interested in pursuing a niche strategy should collect data sensitive to such differences. The evaluation of market creation performance would acknowledge the number of segment-level customers created, given segment size estimates. Our initial framework can expand to include additional input, output, and intermediate variables, as well as control variables, such as those that reflect the diversity of national environments (e.g., subsidiary lifecycle stage, infrastructure availability, literacy rates, cultural factors). Whereas we examine the effects of various strategic marketing levels on market penetration and subsidiary performance within the context of one MNE, further research might examine specific relationships among these variables across a group of MNEs.

Finally, additional research might extend our suggested approach both vertically and laterally. The units being evaluated could be various divisions within a large firm or subsidiary or the entire MNE network. Lateral comparisons between two or more MNEs might focus on the country-level operations of each. Such an investigation could identify why some apparently strong global brands underperform in some markets or how country differences affect MNE performance differences. For example, a MNE that focuses primarily on emerging markets may perform poorly overall compared with another that focuses on developed markets, especially because the former may be focused more on market creation, whereas the latter is attempting to harness its marketing assets. Longitudinal data could clarify the effects of lagged variables.

References

Andersen P, Petersen NC (1993) A procedure for ranking efficient units in data envelopment analysis. Manag Sci 39(October):1261–1264

Anderson JC (1995) Relationships in business markets: exchange episodes, value creation, and their empirical assessment. J Acad Mark Sci 23(Fall):346–350

Banker RD, Hsihui Chang (2006) The super-efficiency procedure for outlier identification, not ranking efficient units. Eur J Oper Res 175(2):1311–1320

Banker RD, Das S, Datar SM (1989) Analysis of cost variances for management control in hospitals. Res Gov Nonprofit Account 5:268–291

Bartlett CA, Ghoshal S (1989) Managing across borders. Harvard University Press, Boston

Basu A, Basu A, Batra R (1995) Modeling the response to direct marketing campaigns. J Mark Res 32(May):204–212

Bharadwaj SG, Rajan Varadajaran P, Fahey J (1993) Sustainable competitive advantage in service industries: a conceptual model and research propositions. J Mark 57(October):83–99

Birkinshaw JM, Morrison AJ (1995) Configurations of strategy and structure in subsidiaries of multinational corporations. J Int Bus Stud 26(4):729–754

Buchanan L, Simmons CJ, Bickart BA (1999) Brand equity dilution: retailer display and context brand effects. J Mark Res 36(August):345–355

Carpano C, Chrisman JJ, Roth K (1994) International strategy and environment: an assessment of the performance relationship. J Int Bus Stud 25(3):639–656

Charnes A, Cooper WW, Rhodes E (1978) Measuring efficiency of decision making units. Eur J Oper Res 2(6):429–444

Cooper WW, Huang ZM, Lelas V, Xi S, Olesen OB (1998) Chance constrained programming formulations for stochastic formulations of efficiency and dominance in DEA. J Prod Anal 9:53–79

Craig CS, Douglas SP (2000) Configural advantage in global markets. J Int Mark 8(1):6–26

Dekimpe MG, Hanssens DM (1995) The persistence of marketing effects on sales. Mark Sci 14 (1):1–21

Donthu N, Hershberger EK, Osmonbekov T (2005) Benchmarking marketing productivity using data envelopment analysis. J Bus Res 58(November):1474–1482

Dowling GR, Uncles M (1997) Do customer loyalty programs really work? Sloan Manage Rev 38 (Summer):71–82

Doyle P (2000) Value-based marketing: marketing strategies for corporate growth and shareholder value. John Wiley & Sons, New York

Frost TS, Birkinshaw JM, Ensign PC (2002) Centers of excellence in multinational corporations. Strateg Manag J 23(November):997–1018

Greene W (2000) Econometric analysis, 4th edn. Prentice-Hall, Englewood Cliffs

Grewal D, Levy M, Mehrotra A, Sharma A (1999) Planning merchandising decisions to account for regional and product assortment differences. J Retail 75(3):405–424

Hewett K, Bearden WO (2001) Dependence, trust, and relational behavior on the part of foreign subsidiary marketing operations: implications for managing global marketing operations. J Mark 65(October):51–66

Kamakura WA, Ramaswami S, Srivastava RK (1991) Applying latent trait analysis in the evaluation of prospects for cross-selling of financial services. Int J Res Mark 8(4):329–349

Kamakura WA, Lenartowicz T, Ratchford BT (1996) Productivity assessment of multiple retail outlets. J Retail 72(4):333–356

Keh HT, Chu S, Xu J (2006) Efficiency, effectiveness and productivity of marketing services. Eur J Oper Res 170:265–276

Kim WC, Mauborgne RA (1993) Effectively conceiving and executing multinationals' worldwide strategies. J Int Bus Stud 24(3):419–448

Kohli AK, Jaworski BJ (1990) Market orientation: the construct, research propositions. J Mark 54 (April):1–18

Luo Y (2002) Capability exploitation and building in a foreign market: implications for multinational enterprises. Organ Sci 13(January–February):48–63

Mahajan J (1991) A data envelopment analytic model for assessing the relative efficiency of the selling function. Eur J Oper Res 53(2):189–205

Murthi BPS, Srinivasan K, Kalyanaram G (1996) Controlling for observed and unobserved managerial skills in determining first-mover market share advantage. J Mark Res 33 (August):329–336

Pilling BK, Donthu N, Henson S (1999) Accounting for the impact of territory characteristics on sales performance: relative efficiency as a measure of salesperson performance. J Pers Sell Sales Manag 19(2):35–45

Porter ME (1986) Changing patterns of international competition. Calif Manage Rev 28 (Winter):9–31

Reinartz W, Kumar V (2000) On the profitability of long lifetime customers: an empirical investigation and implications for marketing. J Mark 64(October):17–35

Reinartz W, Kumar V (2003) On the profitability of long-life customers in a noncontractual setting: an empirical investigation and implications for marketing. J Mark 64(October):17–35

Rust RT, Ambler T, Carpenter GS, Kumar V, Srivastava RK (2004) Measuring marketing productivity: current knowledge and future directions. J Mark 68(October):76–89

Seiford LM, Joe Zhu (1998) An acceptance system decision rule with data envelopment analysis. Comput Oper Res 25(April):329–332

Seiford LM, Joe Zhu (1999) Infeasibility of super efficiency data envelopment analysis models. Inf Syst Oper Res 37(May):174–187

Shapiro AC (2006) Multinational financial management, 8th edn. Wiley, Hoboken

Sheth JN, Sisodia RS (2002) Marketing productivity issues and analysis. J Bus Res 55(5):349–362

Srivastava RK, Shervani TA, Fahey L (1998) Market-based assets and shareholder value: a framework for analysis. J Mark 62(January):2–18

Srivastava RK, Shervani TA, Fahey L (1999) Marketing, business processes, and shareholder value: an organizationally embedded view of marketing activities and the discipline of marketing. J Mark 63(Special Issue):168–179

Syzmanski DM, Bharadwaj SG, Rajan Varadarajan P (1993) Standardization versus adaptation of international marketing strategy: an empirical investigation. J Mark 57(October):1–17

Taggart JH (1997) Autonomy and procedural justice: a framework for evaluating subsidiary strategy. J Int Bus Stud 28(First Quarter):51–76

Tallman SB (1991) Strategic management models and resource-based strategies among MNEs in a host market. Strateg Manag J 12(Summer):69–82

Wind Y, Mahajan V, Swire DJ (1983) An empirical comparison of standardized portfolio models. J Mark 47(Spring):89–99

Yip GS, Madsen TL (1996) Global account management: the new frontier in relationship marketing. Int Mark Rev 13(3):24–42

Zhao H, Luo Y (2002) Product diversification, ownership structure, and subsidiary performance in China's dynamic market. Manag Int Rev 42(First Quarter):27–48

Zhu J (2003) Quantitative model for performance evaluation and benchmarking: data envelopment analysis with spreadsheets. Kluwer Academic, Boston

Zou S, Tamer Cavusgil S (2002) The GMS: a broad conceptualization of global marketing strategy and its effect on firm performance. J Mark 66(October):40–56

Chapter 17
Nonparametric Estimates of the Components of Productivity and Profitability Change in U.S. Agriculture

Christopher J. O'Donnell

Abstract Recent theoretical advances in total factor productivity (TFP) measurement mean that TFP indexes can now be exhaustively decomposed into unambiguous measures of technical change and efficiency change. To date, all applications of this new methodology have involved decomposing indexes that have poor theoretical properties. This article shows how the methodology can be used to decompose a new TFP index that satisfies all economically-relevant axioms from index theory. The application is to state-level data from 1960 to 2004. In most states, the main drivers of agricultural TFP change are found to have been technical change and scale and mix efficiency change.

Keywords Data envelopment analysis • Environmental change • Lowe index • Mix efficiency • Scale efficiency • Technical efficiency • Total factor productivity

It is difficult, if not impossible, to find coherent estimates of the technical change and efficiency change components of U.S. agricultural productivity change. Several estimates of technical change and efficiency change are available (e.g., Morrison Paul and Nehring 2005; Morrison Paul et al. 2004) but they are not coherent in the sense that they do not combine to yield recognizable productivity indexes. And while several researchers have decomposed well-known productivity indexes into various components (e.g., Capalbo 1988), not all of these components have unambiguous interpretations as measures of technical change or efficiency change. This lack of coherent information on the components of productivity change can lead to poor public policy—policy-makers cannot properly assess whether the payoffs from improving the rate of technical progress (e.g., through increased

Reproduction of O'Donnell, C.J., Nonparametric estimates of the components of productivity and profitability change in U.S. agriculture. American Journal of Agricultural Economics, 2012, 94(4):873–890, by permission of Oxford University Press.

C.J. O'Donnell (✉)
The University of Queensland, Centre for Efficiency and Productivity Analysis, Brisbane, QLD 4072, Australia
e-mail: c.odonnell1@uq.edu.au; c.odonnell@economics.uq.edu.au

J. Zhu (ed.), *Data Envelopment Analysis*, International Series in Operations Research & Management Science 238, DOI 10.1007/978-1-4899-7684-0_17

R&D expenditure) are more or less likely to outweigh the payoffs from improving levels of either technical efficiency (e.g., through education and training programs) or scale and mix efficiency (e.g., by using taxes and subsidies to change relative prices). This article fills this information gap by decomposing productivity indexes for U.S. agriculture into an exhaustive set of recognizable measures of technical change, technical efficiency change, and scale and mix efficiency change.

The analysis in this article is conducted within the aggregate quantity-price framework first developed in O'Donnell (2008). In that working paper I show how carefully-defined price and quantity aggregates can be used to decompose profitability change (a measure of value change) into the product of a terms-of-trade (TT) index (a measure of price change) and a *multiplicatively-complete* total factor productivity (TFP) index (a measure of quantity change). I also show that, in theory, any multiplicatively-complete TFP index can be exhaustively decomposed into the product of a measure of technical change and several measures of efficiency change. Important features of this methodology are that it does not depend on restrictive assumptions concerning the production technology, and it does not depend on any assumptions concerning firm behaviour or the level of competition in input or output markets. To date, all empirical applications of the methodology have involved the decomposition of multiplicatively-complete Hicks-Moorsteen TFP indexes (e.g., O'Donnell 2010).

The class of multiplicatively-complete TFP indexes also includes the Laspeyres, Paasche, Fisher and Törnqvist indexes. A problem with these well-known indexes (and the Hicks-Moorsteen index) is that they fail to satisfy a commonsense transitivity axiom. Transitivity guarantees that a direct comparison of two observations (i.e., firms or periods) will yield the same estimate of TFP change as an indirect comparison through a third observation. The usual solution to the transitivity problem involves a geometric averaging procedure due to Elteto and Koves (1964) and Szulc (1964). Unfortunately, although they may be transitive, these so-called EKS indexes fail an identity axiom. The identity axiom guarantees that if outputs and inputs are unchanged then the TFP index will take the value one (i.e., indicate that productivity is also unchanged). This article proposes a new TFP index that satisfies both the transitivity axiom and the identity axiom. It then decomposes the index into technical change and efficiency change components. This article refers to the index as a Lowe TFP index because it can be written as the ratio of two indexes that have been attributed to Lowe (1823).

In practice, decomposing TFP indexes into measures of technical change and efficiency change involves estimating the production frontier. The two main approaches to estimating production frontiers are stochastic frontier analysis (SFA) and data envelopment analysis (DEA). The SFA approach is parametric in the sense that it requires parameterisation of the production frontier and various assumptions concerning the parameters of the distributions of random error terms. One of the advantages of the SFA approach is that it makes an allowance for measurement errors and other sources of statistical noise. Disadvantages of the approach are that results may be sensitive to parametric assumptions, it is difficult to identify the pure scale efficiency change and pure mix efficiency change components of TFP change (i.e., the productivity dividends associated with changes in scale alone and changes in output mix or input mix alone), and results may be unreliable if sample sizes are small. The alternative DEA approach is commonly referred to as a nonparametric approach because it doesn't require any explicit assumptions about the functional form of the frontier (the frontier is implicitly assumed to be locally linear[1]) or the distributions of random error terms (all noise effects are implicitly assumed to be zero). The main advantages of the DEA approach are that there are no statistical issues (e.g., endogeneity) associated with estimating multiple-input multiple-output technologies, and it can be used to estimate levels of, and therefore changes in, pure scale efficiency. This article shows how DEA can also be used to estimate the pure mix efficiency change components of Paasche, Laspeyres, Fisher and Lowe TFP indexes. Disadvantages of the DEA approach are that it does not allow for statistical noise and so cannot distinguish inefficiency from noise, and technical efficiency estimates are upwardly biased in small samples.

The main contributions of this article are threefold. First, it defines a new TFP index (referred to as a Lowe TFP index) that satisfies all economically-relevant axioms from index number theory, including the identity and transitivity axioms. Second, it develops new linear programming methodology for exhaustively decomposing Paasche, Laspeyres, Fisher and Lowe TFP indexes into measures of technical change and various measures of technical, scale and mix efficiency change. Third, it fills an information gap by reporting a coherent set of estimates of productivity change, technical change and efficiency change in US agriculture.

The structure of the article is as follows. The first few sections explain how profitability change can be decomposed into the product of a TFP index and a TT index. Attention is focused on the Lowe TFP index because this is one of only handful of indexes that satisfy seven basic axioms from index number theory. Some time is spent explaining how the methodology I develop in O'Donnell (2008) can be used to decompose Lowe TFP indexes into measures of technical change and efficiency change. This involves the specification of new DEA problems for estimating levels of (and therefore changes in) pure technical, scale and mix efficiency. Variants of these DEA problems can also be used to estimate the maximum level of TFP possible using a production technology. The last few sections of the article describe the empirical application. The dataset is a state-level panel dataset assembled by the Economic Research Service (ERS) of the U.S. Department of Agriculture (USDA). It comprises observations on the prices and quantities of agricultural outputs and inputs in $I = 48$ states over the $T = 45$ years from 1960 to 2004. The penultimate section reports estimates of TFP change and various measures of efficiency change for selected states in selected periods. The final section of the article summarises the findings and offers two suggestions for further research.

[1] Local linearity means the frontier is formed by a number of intersecting hyperplanes. Thus, it may be more appropriate to refer to DEA as a semiparametric rather than a nonparametric approach.

17.1 The Components of Profitability Change

Let $\boldsymbol{x}_{it} \in \Re_{+}^{M}$, $\boldsymbol{q}_{it} \in \Re_{+}^{N}$, $\boldsymbol{w}_{it} \in \Re_{+}^{M}$ and $\boldsymbol{p}_{it} \in \Re_{+}^{N}$ denote vectors of input and output quantities and prices for firm i in period t. In O'Donnell (2008) I define the TFP of the firm to be $TFP_{it} = Q_{it}/X_{it}$ where $Q_{it} \equiv Q(\boldsymbol{q}_{it})$ is an aggregate output and $X_{it} \equiv X(\boldsymbol{x}_{it})$ is an aggregate input. The only requirements placed on the aggregator functions $Q(.)$ and $X(.)$ are that they be nonnegative, nondecreasing and linearly homogeneous. If TFP is defined in this way then the index that compares the TFP of firm i in period t with the TFP of firm h in period s is

$$TFPI_{hsit} = \frac{TFP_{it}}{TFP_{hs}} = \frac{Q_{it}/X_{it}}{Q_{hs}/X_{hs}} = \frac{Q_{it}/Q_{hs}}{X_{it}/X_{hs}} = \frac{QI_{hsit}}{XI_{hsit}} \tag{17.1}$$

where $QI_{hsit} = Q_{it}/Q_{hs}$ and $XI_{hsit} = X_{it}/X_{hs}$ are output and input quantity indexes. Thus, within this framework, TFP growth is a measure of output growth divided by a measure of input growth, which is how productivity growth is usually defined (e.g., Griliches 1961; Jorgenson and Griliches 1967). In O'Donnell (2008) I use the term *multiplicatively-complete* to refer to TFP indexes that can be written in terms of aggregate input and output quantities as in (17.1). The Laspeyres, Paasche, Törnqvist, Hicks-Moorsteen and Fisher TFP indexes are all multiplicatively complete. Indexes that are *not* multiplicatively-complete include the widely-used output- and input-oriented Malmquist TFP indexes of Caves et al. (1982) and the lesser-known TFP index of Diewert and Morrison (1986)—except in restrictive special cases, these indexes cannot be expressed as an output quantity index divided by an input quantity index (i.e., they cannot be used to measure TFP change).

Associated with any non-zero aggregate quantities are implicit aggregate prices $P_{it} = \boldsymbol{p}_{it}^{\prime}\boldsymbol{q}_{it}/Q_{it}$ and $W_{it} = \boldsymbol{w}_{it}^{\prime}\boldsymbol{x}_{it}/X_{it}$. The existence of these implicit prices means that profit can be written $\pi_{it} = P_{it}Q_{it} - W_{it}X_{it}$ and profitability can be written $PROF_{it} = (P_{it}Q_{it})/(W_{it}X_{it})$. Furthermore, the index that compares the profitability of firm i in period t with the profitability of firm h in period s can be written

$$PROFI_{hsit} = \frac{PROF_{it}}{PROF_{hs}} = \frac{P_{it}Q_{it}}{W_{it}X_{it}} \times \frac{W_{hs}X_{hs}}{P_{hs}Q_{hs}} = \frac{PI_{hsit}}{WI_{hsit}} \times \frac{QI_{hsit}}{XI_{hsit}} = TTI_{hsit} \times TFPI_{hsit} \tag{17.2}$$

where $PI_{hsit} = P_{it}/P_{hs}$ is an output price index, $WI_{hsit} = W_{it}/W_{hs}$ is an input price index and $TTI_{hsit} = PI_{hsit}/WI_{hsit}$ is a TT index measuring output price change relative to input price change. It is apparent from (17.2) that (1) if the reference and comparison firms receive the same prices for their outputs and pay the same prices for their inputs then the TT index will equal unity and any changes in profitability will be plausibly attributed entirely to changes in TFP, (2) if two firms use the same inputs to produce the same outputs then any changes in profitability will be attributed entirely to changes in prices, and (3) if profitability is constant then a TFP index can be computed as the reciprocal of a TT index.

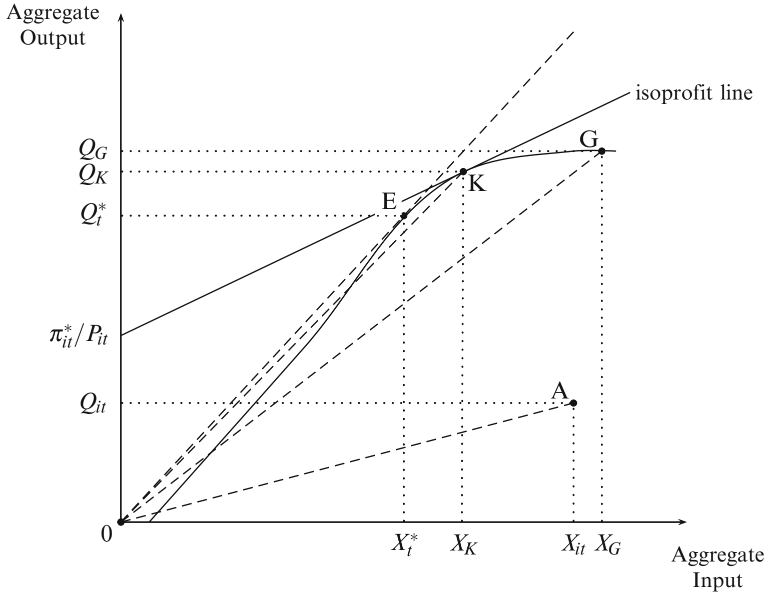


Fig. 17.1 TFP, profitability and the TT

This aggregate quantity-price framework can be used to provide important insights into the behaviour of rational profit-maximising firms. To illustrate, Fig. 17.1 depicts the aggregate output and input of firm i in period t in two dimensional-aggregate quantity space (point A). In this figure, the curve passing through points E, K and G is the boundary of the set of all aggregate-output aggregate-input combinations that are technically feasible in period t. The definition (17.1) means that the TFP of a firm operating at any point in aggregate quantity space is the slope of the ray from the origin to that point—for example, the TFP of the firm operating at point A is $TFP_{it} = Q_{it}/X_{it} =$ slope 0A, and the TFP of the firm operating at point E is $TFP_t^* = Q_t^*/X_t^* =$ slope 0E. The solid line passing through point K in Fig. 17.1 is an isoprofit line with slope $-W_{it}/P_{it}$ and intercept π_{it}^*/P_{it}. The fact that this isoprofit line is tangent to the production frontier means that point K maximizes profit at aggregate prices P_{it} and W_{it}. For the technology represented in Fig. 17.1 (there are other technologies where this may not be true), the point of maximum profit will coincide with the point of maximum TFP if and only if the level of maximum TFP (the slope of the ray 0E) equals the reciprocal of the TT (the slope of the isoprofit line). This equality between the level of maximum TFP and the reciprocal of the TT is a characteristic of perfectly competitive markets and, in such cases, profits are zero. It is clear that a rational efficient firm having a benefit function that is increasing in net returns will be drawn away from the point of maximum TFP in response to an improvement in its TT, to a point such as K or G. The associated inequality between the TT and the level of maximum TFP is a characteristic of non-competitive markets and, in such cases, maximum profits are strictly non-zero. Point G in Fig. 17.1 is the profit maximising solution in the limiting case where all inputs are relatively costless. For rational efficient firms, the economically feasible region of production is the region of locally-decreasing returns to scale between points E and G. Productivity falls and profits rise as rational efficient firms move optimally from point E to point G. Conversely, productivity increases and profits fall as rational efficient firms move optimally from point G to point E in response to deteriorations in their terms-of-trade.

This inverse relationship between TFP and the TT has two interesting implications. First, it provides a rationale for microeconomic reform programs designed to increase levels of competition in agricultural output and input markets—deteriorations in the TT that result from increased competition will tend to drive firms towards points of maximum TFP. Second, it may provide an explanation for convergence in rates of agricultural TFP growth in regions, states and countries that are becoming increasingly integrated and/or globalized—firms that strictly prefer more income to less and who face the same technology and prices will optimally choose to operate at the same point on the production frontier, they will make similar adjustments to their production choices in response to changes in the common TT, and they will thus experience similar rates of TFP change.

An empirical illustration of the inverse relationship between agricultural TFP and the agricultural TT is provided in Fig. 17.2. The solid lines (with or without markers) in this figure depict changes in profitability ($\Delta PROF$), total factor productivity (ΔTFP) and the terms of trade (ΔTT) in Alabama over the period 1960–2004 (Alabama 1960 = 1). The TFP index (a Lowe index) indicates that TFP

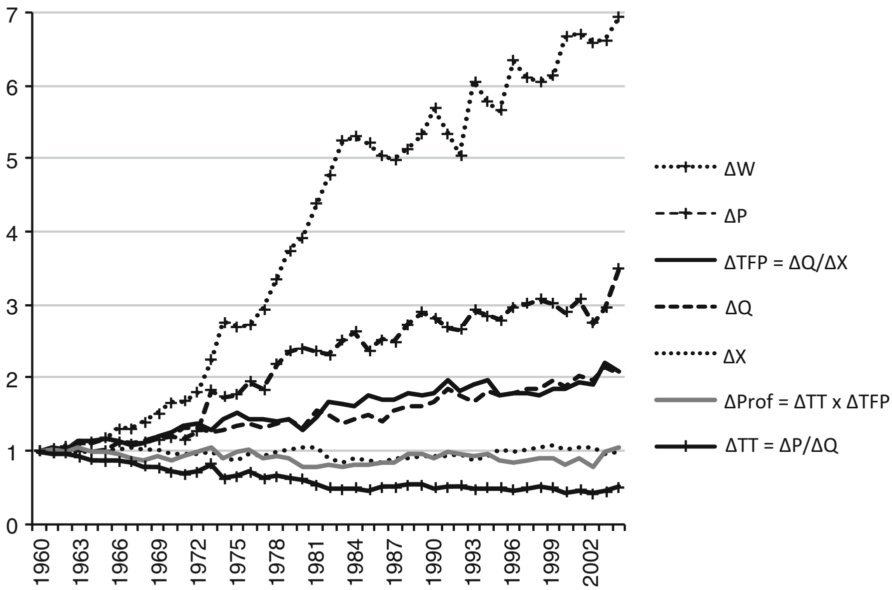


Fig. 17.2 Profitability, quantity and price change in Alabama (1960 = 1)

has more than doubled over this period. However, profitability has increased by only 4.6 % due to the offsetting effects of an estimated 49.6 % deterioration in the TT (i.e., $\Delta PROF = \Delta TFP \times \Delta TT = 2.076 \times 0.504 = 1.046$). The dashed and dotted lines without markers depict the output change (ΔQ) and input change (ΔX) components of TFP change—these series reveal that virtually all TFP growth in Alabama over the sample period has been due to output growth (i.e., $\Delta TFP = \Delta Q/\Delta X = 2.066/0.995 = 2.076$). The dashed and dotted lines with markers depict the associated output and input price indexes (ΔP and ΔW)—these series reveal that input prices have increased seven-fold over the sample period but output prices have increased by only half that amount (i.e., $\Delta TT = \Delta P/\Delta W = 3.493/6.936 = 0.504$). The output and input price and quantity indexes presented in Fig. 17.2 can be further decomposed into indexes that measure changes in the prices and quantities of a disaggregated set of outputs and inputs. For example, the measured increase in output prices over the sample period can be attributed to a 294 % increase in livestock prices, a 151 % increase in crop prices, and a 381 % increase in the prices of other agricultural outputs. These price increases have been associated with a 158 % increase in livestock outputs, a 30 % increase in crop outputs and a 141 % increase in other agricultural outputs. Further analysis of relationships between output and input prices and quantities at this level of disaggregation is straightforward (e.g., O'Donnell et al. 1999) but beyond the scope of this article.

17.2 Lowe TFP Indexes

In O'Donnell (2008) I explicitly identify the aggregator functions that underpin Laspeyres, Paasche, Fisher, Törnqvist and Hicks-Moorsteen price, quantity and TFP indexes. Unfortunately, these indexes are not generally suitable for making multitemporal (i.e., many period) or multilateral (i.e., many firm) comparisons of TFP because they violate at least one important axiom from index number theory. This article uses a very simple linear aggregator function to derive TFP indexes that satisfy all basic index number axioms. Specifically, the article aggregates outputs and inputs using the functions $Q(\boldsymbol{q}_{it}) \propto \boldsymbol{p}_0'\boldsymbol{q}_{it}$ and $X(\boldsymbol{x}_{it}) \propto \boldsymbol{w}_0'\boldsymbol{x}_{it}$ where $\boldsymbol{p}_0$ and $\boldsymbol{w}_0$ are pre-determined firm- and time-invariant reference prices. The associated output quantity, input quantity and TFP indexes that compare firm i in period t with firm h in period s are

$$QI_{hsit} = \frac{Q(\boldsymbol{q}_{it})}{Q(\boldsymbol{q}_{hs})} = \frac{\boldsymbol{p}_0'\boldsymbol{q}_{it}}{\boldsymbol{p}_0'\boldsymbol{q}_{hs}}, \tag{17.3}$$

$$XI_{hsit} = \frac{X(\boldsymbol{x}_{it})}{X(\boldsymbol{x}_{hs})} = \frac{\boldsymbol{w}_0'\boldsymbol{x}_{it}}{\boldsymbol{w}_0'\boldsymbol{x}_{hs}} \quad \text{and} \tag{17.4}$$

$$TFPI_{hsit} = \frac{QI_{hsit}}{XI_{hsit}} = \frac{\boldsymbol{p}_0'\boldsymbol{q}_{it}}{\boldsymbol{p}_0'\boldsymbol{q}_{hs}}\,\frac{\boldsymbol{w}_0'\boldsymbol{x}_{hs}}{\boldsymbol{w}_0'\boldsymbol{x}_{it}}. \tag{17.5}$$

These indices are ratios of the values of different baskets of goods evaluated at the same set of reference prices. This article refers to the TFP index (17.5) as a Lowe index because the component output and input quantity indexes given by (17.3) and (17.4) are commonly attributed to Lowe (1823) (e.g., Hill 2008).

Any number of price vectors can be used as reference price vectors $\boldsymbol{p}_0$ and $\boldsymbol{w}_0$ in the indexes (17.3)–(17.5). This article recommends using price vectors that are representative of the price vectors faced by all firms that are to be compared (usually all firms in the dataset).[2] In practice, statistical tests can be used to assess whether a chosen vector of reference prices is representative of the prices faced by any given set of firms. For example, the reference prices used for the empirical work in this article are the sample means of prices in all states in all periods (i.e., $\boldsymbol{p}_0 = \bar{p} = (IT)^{-1}\sum_i\sum_t \boldsymbol{p}_{it}$ and $\boldsymbol{w}_0 = \bar{w} = (IT)^{-1}\sum_i\sum_t \boldsymbol{w}_{it}$). If observations for several more years were to become available then a simple Wald test[3] could be used to determine whether this same vector of reference prices was representative of prices in the larger sample.

Lowe indices satisfy a number of important axioms. To avoid repetition, this article only presents the axioms satisfied by the Lowe output quantity index (17.3). Analogous axioms are satisfied by the input quantity index (17.4) and the TFP index (17.5).

The Lowe output quantity index (17.3) is a function $QI(\boldsymbol{q}_{hs}, \boldsymbol{q}_{it}, \boldsymbol{p}_0)$ that satisfies the following commonsense axioms:

A1 Monotonicity[4]: $QI(\boldsymbol{q}_{hs}, \boldsymbol{q}_{jr}, \boldsymbol{p}_0) > QI(\boldsymbol{q}_{hs}, \boldsymbol{q}_{it}, \boldsymbol{p}_0)$ if $\boldsymbol{q}_{jr} \geq \boldsymbol{q}_{it}$ and $QI(\boldsymbol{q}_{jr}, \boldsymbol{q}_{it}, \boldsymbol{p}_0) < QI(\boldsymbol{q}_{hs}, \boldsymbol{q}_{it}, \boldsymbol{p}_0)$ if $\boldsymbol{q}_{jr} \geq \boldsymbol{q}_{hs}$;
A2 Linear homogeneity: $QI(\boldsymbol{q}_{hs}, \lambda\boldsymbol{q}_{it}, \boldsymbol{p}_0) = \lambda QI(\boldsymbol{q}_{hs}, \boldsymbol{q}_{it}, \boldsymbol{p}_0)$ for $\lambda > 0$;
A3 Identity: $QI(\boldsymbol{q}_{it}, \boldsymbol{q}_{it}, \boldsymbol{p}_0) = 1$;
A4 Homogeneity of degree zero: $QI(\lambda\boldsymbol{q}_{hs}, \lambda\boldsymbol{q}_{it}, \boldsymbol{p}_0) = QI(\boldsymbol{q}_{hs}, \boldsymbol{q}_{it}, \boldsymbol{p}_0)$ for $\lambda > 0$;
A5 Commensurability: $QI(\boldsymbol{q}_{hs}\boldsymbol{\Lambda}, \boldsymbol{q}_{it}\boldsymbol{\Lambda}, \boldsymbol{p}_0\Lambda^{-1}) = QI(\boldsymbol{q}_{hs}, \boldsymbol{q}_{it}, \boldsymbol{p}_0)$ where $\boldsymbol{\Lambda}$ is a diagonal matrix with diagonal elements strictly greater than zero;
A6 Proportionality: $QI(\boldsymbol{q}_{hs}, \lambda\boldsymbol{q}_{hs}, \boldsymbol{p}_0) = \lambda$ for $\lambda > 0$; and
A7 Transitivity: $QI_{hsit} = QI_{hsjr}QI_{jrit}$.

Axiom A1 (monotonicity) means that the index increases with increases in any element of the comparison vector $\boldsymbol{q}_{it}$ and/or with decreases in any element of the reference vector $\boldsymbol{q}_{hs}$. Axiom A2 (linear homogeneity) means that a proportionate

[2] More precisely, reference prices should be representative of the relative importance (i.e., relative value) that decision-makers place on different outputs and inputs. Observed prices are not always the best measures of relative importance.

[3] For example, the prices in the larger sample could be used to test the joint null hypothesis that the population mean prices are equal to the reference prices. Such a test can be conducted in a regression framework using test commands available in standard econometrics software packages.

[4] Let q_{nit} denote the nth element of $\boldsymbol{q}_{it}$. The notation $\boldsymbol{q}_{hs} \geq \boldsymbol{q}_{it}$ means that $q_{nhs} \geq q_{nit}$ for $n = 1, \ldots, N$ and there exists at least one value $n \in \{1, \ldots, N\}$ where $q_{nhs} > q_{nit}$.

increase in the comparison vector will cause the same proportionate increase in the index. Axiom A3 (identity) means that if the comparison and reference vectors are identical then the index number takes the value one. Axiom A4 (homogeneity of degree zero) means that multiplication of the comparison and reference vectors by the same constant will leave the index number unchanged. Axiom A5 (commensurability) means that a change in the units of measurement of an output (e.g., from kilograms to tonnes) does not change the value of the index. Axiom A6 (proportionality) means that if the comparison vector is proportionate to the reference vector then the index number is equal to the factor of proportionality. Finally, A7 says the index number that directly compares the outputs of a comparison firm/period with the outputs of a reference firm/period is identical to the index number computed when the comparison is made through an intermediate firm/period.

The Laspeyres, Paasche, Fisher, Törnqvist, Hicks-Moorsteen and Lowe indexes all satisfy axioms A1–A6, but, of these indexes, only the Lowe index also satisfies the transitivity axiom A7. In practice, it is common to compute intransitive Fisher and Törnqvist indices and then address the transitivity problem by applying a geometric averaging procedure proposed by Elteto and Koves (1964) and Szulc (1964). Unfortunately, index numbers constructed using this method fail the identity axiom A3. This means that EKS indexes will almost certainly indicate increases or decreases in TFP even when input-output combinations (i.e., levels of TFP) haven't changed.

The focus of this article is first and foremost on the measurement of quantity change and TFP change. Lowe indexes are ideal for this purpose because they satisfy axioms A1–A7. However, is worth noting that Lowe implicit indexes are less than ideal for the purpose of measuring changes in prices and the TT. For example, the implicit Lowe output price index $PI(\boldsymbol{q}_{hs},\boldsymbol{q}_{it},\boldsymbol{p}_{hs},\boldsymbol{p}_{it},\boldsymbol{p}_0) \equiv (\boldsymbol{p}_{it}'\boldsymbol{q}_{it})/\boldsymbol{p}_{hs}'\boldsymbol{q}_{hs})/(\boldsymbol{p}_0'\boldsymbol{q}_{it})/\boldsymbol{p}_0'\boldsymbol{q}_{hs})$ does not generally[5] satisfy price analogues of the identity and proportionality axioms A3 and A6. Implicit Laspeyres, Paasche, Fisher, Törnqvist, Hicks-Moorsteen and EKS-type price and TT indexes are also less than ideal because they also fail to satisfy at least one axiom.

To illustrate the importance of selecting a theoretically-plausible TFP index formula, Fig. 17.3 presents (implausible) EKS and (plausible) Lowe indexes of agricultural TFP change in Alabama, Florida and Wyoming over the period 1960–2004 (Alabama 1960 = 1). The grey (solid, dashed and dotted) lines in this figure are indexes computed by applying the EKS procedure to binary Fisher indexes, and the black lines are Lowe indexes defined by (17.5) (the Lowe index for Alabama was depicted earlier in Fig. 17.2). Observe that there are important differences between the two sets of estimates for particular states in particular years. For example, the EKS index indicates that in 1967 Alabama and Wyoming

[5] For this index, the identity axiom requires $PI(\boldsymbol{q}_{hs},\boldsymbol{q}_{it},\boldsymbol{p}_{it},\boldsymbol{p}_{it},\boldsymbol{p}_0) = 1$ while the proportionality axiom requires $PI(\boldsymbol{q}_{hs},\boldsymbol{q}_{it},\boldsymbol{p}_{hs},\lambda\boldsymbol{p}_{hs},\boldsymbol{p}_0) = \lambda$ for $\lambda > 0$. Both axioms will be satisfied if $\boldsymbol{p}_0 \propto \boldsymbol{p}_{hs} \propto \boldsymbol{p}_{it}$ (e.g., if there is no price change in the dataset and $\boldsymbol{p}_0 = \bar{\boldsymbol{p}}$; or if $\boldsymbol{p}_{hs} \propto \boldsymbol{p}_{it}$ for all $h = 1, \ldots, I$ and $s = 1, \ldots, T$ and $\boldsymbol{p}_0 = \bar{\boldsymbol{p}}$).

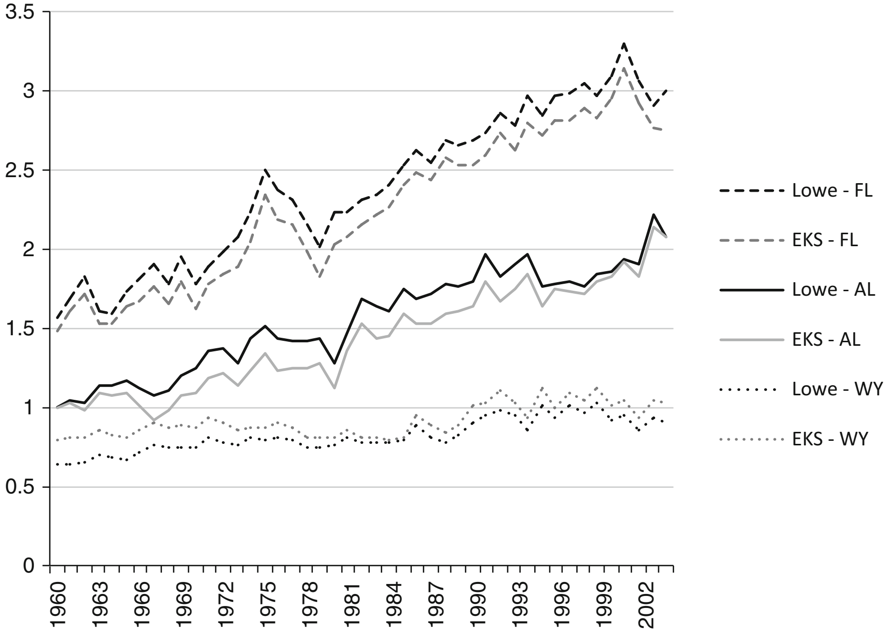


Fig. 17.3 TFP change in Florida, Alabama and Wyoming (Alabama 1960 = 1)

were equally productive, but the Lowe index indicates that Alabama was 39 % more productive than Wyoming.

Explaining changes in TFP over time and space involves estimating measures of technical change and efficiency change. The following section describes how multiplicatively-complete TFP indexes such as the Lowe index can be exhaustively decomposed into such measures. The EKS index cannot generally be decomposed this way because it is not multiplicatively-complete.

17.3 The Components of TFP Change

In O'Donnell (2008) I show that any multiplicatively-complete TFP index can be decomposed into measures of technical change and efficiency change. Among the efficiency change components are input- and output-oriented measures of technical, scale and mix efficiency change. The decomposition methodology involves identifying points of economic interest in aggregate quantity space. For illustrative purposes, consider a two-output technology and let $\boldsymbol{q}_{it} = (q_{1it}, q_{2it})'$ denote the vector of outputs produced by firm i in period t. Let $\boldsymbol{p}_{it} = (p_{1it}, p_{2it})'$ be the associated vector of output prices and, for a simple exposition that is consistent with the Lowe indexes presented in Fig. 17.3, let outputs be aggregated using the

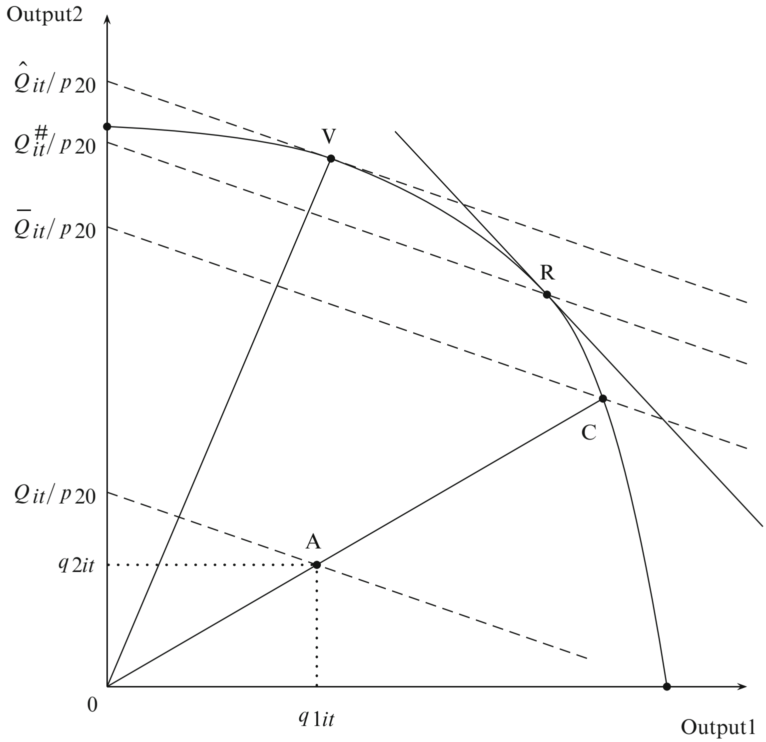


Fig. 17.4 Output-oriented measures of efficiency ($N = 2$)

Lowe output aggregator function $Q(\boldsymbol{q}_{it}) = \boldsymbol{p}_0'\boldsymbol{q}_{it}$ where $\boldsymbol{p}_0 = (p_{10}, p_{20})'$. Figure 17.4 depicts measures of technical, mix and revenue-allocative efficiency for this firm in output space: the curve passing through points V, R and C is the familiar production possibilities frontier; the solid line that is tangent to the frontier at point R is an isorevenue line with slope $-p_{1it}/p_{2it}$; and the dashed line passing through point A is an iso-aggregate-output line with slope $-p_{10}/p_{20}$ and intercept Q_{it}/p_{20}. For the firm producing $\boldsymbol{q}_{it}$, maximising output while holding the output mix fixed involves a move from point A to point C and an increase in the aggregate output from Q_{it} to $\bar{Q}_{it}$, maximising revenue without any restrictions on the output mix involves a move to point R and an increase in the aggregate output to $Q^{\#}_{it}$, and maximising aggregate output without any restrictions on the output mix involves a move to point V and an increase in the aggregate output to $\hat{Q}_{it}$. Associated measures of efficiency are: the Farrell (1957) output-oriented measure of technical efficiency, $OTE_{it} = Q_{it}/\bar{Q}_{it}$; the conventional measure of revenue-allocative efficiency, $RAE_{it} = \bar{Q}_{it}/Q^{\#}_{it}$; and a measure of output-oriented mix efficiency first defined in O'Donnell (2008), $OME_{it} = \bar{Q}_{it}/\hat{Q}_{it}$.

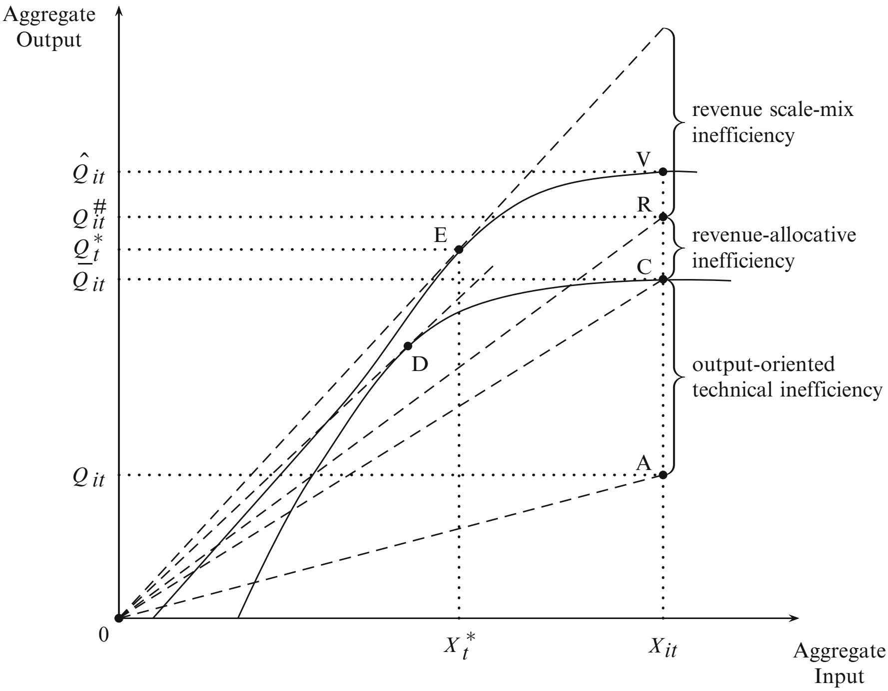


Fig. 17.5 Output-oriented measures of efficiency ($N \geq 1$)

Figure 17.5 maps the points A, C, R and V from Fig. 17.4 into aggregate quantity space. In this figure, the curve passing through points C and D is a *mix-restricted* frontier—this frontier envelops all technically-feasible aggregate input-output combinations that have the same input mix and output mix as the firm operating at point A. The curve passing through points V and E is the *unrestricted* production frontier depicted earlier in Fig. 17.1—this frontier envelops all aggregate input-output combinations that are feasible when all mix restrictions are relaxed. Recall that the maximum TFP that is possible using this technology is the TFP at point E: $TFP_t^* = Q_t^*/X_t^* =$ slope OE. In O'Donnell (2008) I define TFP efficiency to be the difference between observed TFP and this maximum level of TFP:

$$TFPE_{it} = \frac{TFP_{it}}{TFP_t^*} = \frac{\text{slope OA}}{\text{slope OE}}. \tag{17.6}$$

It is clear from Fig. 17.5 that TFP efficiency can be decomposed as:

$$\begin{aligned} TFPE_{it} &= \frac{\text{slope OA}}{\text{slope OE}} = \frac{\text{slope OA}}{\text{slope OC}} \times \frac{\text{slope OC}}{\text{slope OR}} \times \frac{\text{slope OR}}{\text{slope OE}} \\ &= OTE_{it} \times RAE_{it} \times RSME_{it} \end{aligned} \tag{17.7}$$

where $RSME_{it} =$ slope OR/slope OE denotes revenue-scale-mix efficiency (a measure of the difference between TFP at a revenue-allocatively efficient point and TFP at the point of maximum TFP). Other economically-meaningful decompositions of TFP efficiency include O'Donnell (2008):

$$TFPE_{it} = OTE_{it} \times OSE_{it} \times RME_{it} \qquad \text{and} \tag{17.8}$$

$$TFPE_{it} = OTE_{it} \times OME_{it} \times ROSE_{it} \tag{17.9}$$

where $OSE_{it} =$ slope OC/slope OD is the conventional measure of output-oriented scale efficiency (a measure of the difference between TFP at a technically efficient point and the maximum TFP that is possible when holding the output and input mixes fixed), $ROSE_{it} =$ slope OV/slope OE is residual output-oriented scale efficiency (a measure of the difference between TFP at an output-mix-efficient point and the maximum possible TFP), and $RME_{it} =$ slope OD/slope OE is residual mix efficiency (a measure of the difference between TFP at a scale-efficient point and the maximum possible TFP).

Multiplicatively-complete TFP indexes can be conveniently decomposed by rearranging Eq. (17.6) as $TFP_{it} = TFP_t^* \times TFPE_{it}$. A similar equation holds for firm h in period s. Thus, the TFP index defined by (17.1) can be written

$$TFPI_{hsit} = \frac{TFP_{it}}{TFP_{hs}} = \left(\frac{TFP_t^*}{TFP_s^*}\right)\left(\frac{TFPE_{it}}{TFPE_{hs}}\right). \tag{17.10}$$

The first term in parentheses is a measure of the difference in the maximum TFP possible in the two periods—this is a natural measure of technical change. Equation (17.10) reveals that TFP change can be exhaustively decomposed into a measure of technical change and a measure of efficiency change. Equations such as (17.7)–(17.9) can be used to further decompose the efficiency change component into any number of meaningful measures. For example, any multiplicatively-complete TFP index can be decomposed into measures of technical change, technical efficiency change, and a combined measure of scale and mix efficiency change:

$$TFPI_{hsit} = \frac{TFP_{it}}{TFP_{hs}} = \left(\frac{TFP_t^*}{TFP_s^*}\right)\left(\frac{OTE_{it}}{OTE_{hs}}\right)\left(\frac{OSME_{it}}{OSME_{hs}}\right) \tag{17.11}$$

where $OSME_{it} = OSE_{it} \times RME_{it}$ denotes output-oriented scale-mix efficiency (a move from point C to point E in Fig. 17.5).

An empirical illustration of these decompositions is presented in Fig. 17.6. This figure decomposes the Lowe TFP index for Alabama (depicted earlier in Figs. 17.2 and 17.3) into an estimate of technical change (ΔTFP^*) and an estimate of efficiency change ($\Delta TFPE$) (the DEA methodology used to obtain these estimates is discussed in the next section). Observe that TFP in Alabama increased by 107.6 % over the sample period due to the combined effects of an 81.4 % increase

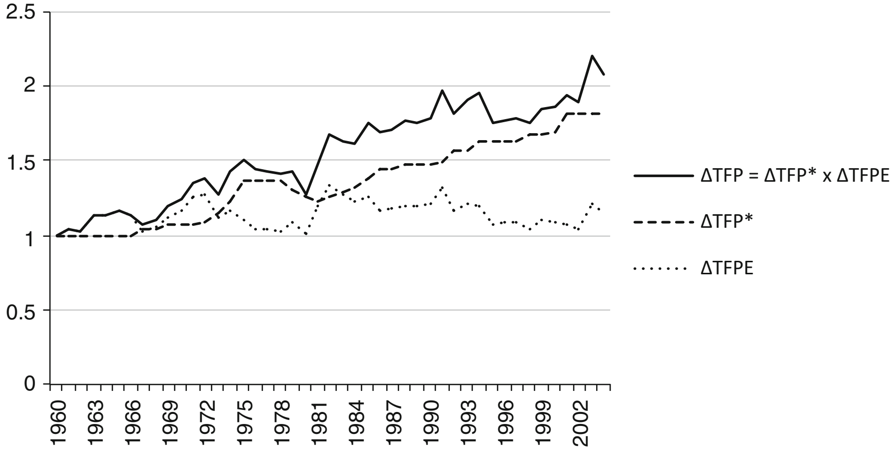


Fig. 17.6 Components of TFP change in Alabama (1960 = 1)

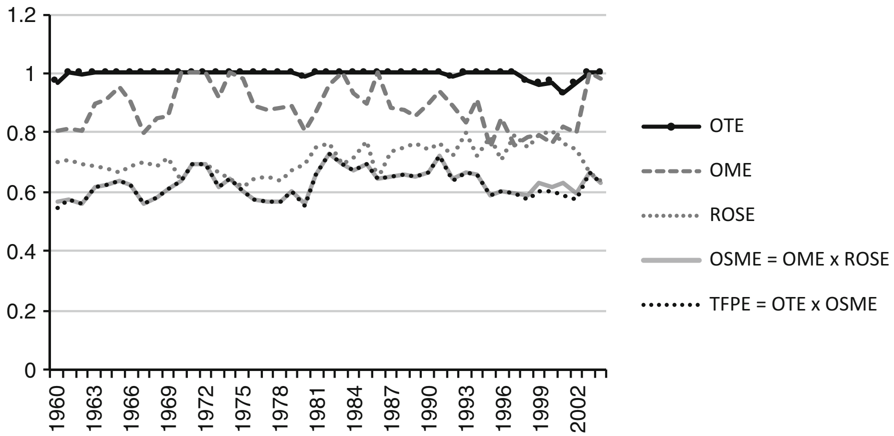


Fig. 17.7 Efficiency levels in Alabama

in the maximum TFP possible (i.e., technical change) and a 14.5 % increase in overall efficiency (i.e., $\Delta TFP = \Delta TFP^* \times \Delta TFPE = 1.814 \times 1.145 = 2.076$). A further breakdown of the efficiency change component is presented in Fig. 17.7. Observe that the level of TFP efficiency in Alabama increased from 0.549 in 1960 to 0.629 in 2004 (implying $\Delta TFPE = 0.629/0.549 = 1.145$). Also observe that Alabama was technically efficient but scale-mix inefficient throughout much of the sample period: in 1960 Alabama was 96.9 % technically efficient but only 56.7 % scale-mix efficient (i.e., $TFPE = OTE \times OSME = 0.969 \times 0.567 = 0.549$); in 2004 the state was fully technically efficient but still only 62.9 % scale-mix efficient (i.e., $TFPE = OTE \times OSME = 1 \times 0.629 = 0.629$). Note that the levels of efficiency depicted in Fig. 17.7 can also be viewed as indexes that compare efficiency levels in Alabama with efficiency levels in any state that was fully efficient in 1960 (e.g., Kansas).

17.4 Estimating Technical, Scale and Mix Efficiency Using DEA

The usual menu of DEA and SFA models and estimators is available for estimating the production frontiers depicted in Figs. 17.4 and 17.5. In O'Donnell (2010) I show how DEA can be used to estimate measures of technical, scale and mix efficiency associated with a Hicks-Moorsteen TFP index. This article extends that DEA methodology to the estimation of measures of efficiency associated with Paasche, Laspeyres, Fisher and Lowe TFP indexes.

A useful starting point is the well-known DEA problem for estimating the output-oriented technical efficiency of firm i in period t. If the technology exhibits variable returns to scale then an estimate of $OTE_{it} = Q_{it}/\bar{Q}_{it}$ can be obtained by solving

$$Q_{it}/\bar{Q}_{it} = \min_{\lambda,\boldsymbol{\theta}}\{\lambda^{-1} : \lambda \boldsymbol{q}_{it} \leq \boldsymbol{Q\theta}; \boldsymbol{X\theta} \leq \boldsymbol{x}_{it}; \boldsymbol{\theta}'\boldsymbol{\iota} = 1; \lambda, \boldsymbol{\theta} \geq 0\} \tag{17.12}$$

where $\boldsymbol{Q}$ is an $N \times J_t$ matrix of observed outputs, $\boldsymbol{X}$ is an $M \times J_t$ matrix of observed inputs, $\boldsymbol{\theta}$ is a $J_t \times 1$ vector, $\boldsymbol{\iota}$ is a $J_t \times 1$ unit vector, and J_t is the number of observations used to estimate the frontier in period t. If the constraint $\boldsymbol{\theta}'\boldsymbol{\iota} = 1$ is deleted from the problem then the estimated technology will exhibit constant returns to scale. Estimates of output-oriented scale efficiency are computed by taking the ratio of the technical efficiency scores estimated under these alternative returns to scale assumptions.

To compute a measure of output-oriented mix efficiency it is convenient to first write the linear program (17.12) in the following equivalent form:

$$Q_{it}/\bar{Q}_{it} = \min_{\lambda,\boldsymbol{\theta},\boldsymbol{z}}\{Q(\boldsymbol{q}_{it})/Q(\boldsymbol{z}) : \boldsymbol{z} \leq \boldsymbol{Q\theta}; \boldsymbol{X\theta} \leq \boldsymbol{x}_{it}; \boldsymbol{\theta}'\boldsymbol{\iota} = 1; \boldsymbol{z} = \lambda \boldsymbol{q}_{it}; \lambda, \boldsymbol{\theta}, \boldsymbol{z} \geq 0\} \tag{17.13}$$

where $Q(.)$ is *any* valid output aggregator function. The equivalence of problems (17.12) and (17.13) can be established by substituting the constraint $\boldsymbol{z} = \lambda \boldsymbol{q}_{it}$ into to the objective function in (17.13) and noting that, since all valid aggregator functions are linearly homogeneous, $Q(\boldsymbol{q}_{it})/Q(\lambda \boldsymbol{q}_{it}) = \lambda^{-1}$. Writing (17.12) in the form of (17.13) is useful because the constraint $\boldsymbol{z} = \lambda \boldsymbol{q}_{it}$ makes it explicit that estimating output-oriented technical efficiency involves estimating the maximum increase in TFP (or aggregate output) that is possible while holding the input level (and therefore the aggregate input) and the output mix fixed. Moreover, it suggests that an estimate of the output-oriented mix efficiency of firm i in period t can be obtained by simply relaxing the mix constraint $\boldsymbol{z} = \lambda \boldsymbol{q}_{it}$. Specifically, an estimate of $OME_{it} = \bar{Q}_{it}/\hat{Q}_{it} = (Q_{it}/\hat{Q}_{it})/(Q_{it}/\bar{Q}_{it})$ can be obtained by solving

$$Q_{it}/\hat{Q}_{it} = \min_{\boldsymbol{\theta},\boldsymbol{z}}\{Q(\boldsymbol{q}_{it})/Q(\boldsymbol{z}) : \boldsymbol{z} \leq \boldsymbol{Q\theta}; \boldsymbol{X\theta} \leq \boldsymbol{x}_{it}; \boldsymbol{\theta}'\boldsymbol{\iota} = 1; \boldsymbol{\theta}, \boldsymbol{z} \geq 0\}. \tag{17.14}$$

Unlike the solution to the technical efficiency problem (17.12), the solution to the mix efficiency problem (17.14) depends on the choice of aggregator function. If interest lies in identifying the output-oriented mix efficiency change component of a particular multiplicatively-complete TFP index, then the aggregator function should be the function that underpins that index. For example, if the TFP index is a Paasche index then the aggregator function is $Q(\boldsymbol{q}_{it}) \propto \boldsymbol{p}_{it}'\boldsymbol{q}_{it}$ and problem (17.14) becomes

$$Q_{it}/\hat{Q}_{it} = \min_{\boldsymbol{\theta},\boldsymbol{z}}\{\boldsymbol{p}_{it}'\boldsymbol{q}_{it}/\boldsymbol{p}_{it}'\boldsymbol{z} : \boldsymbol{z} \leq Q\boldsymbol{\theta}; X\boldsymbol{\theta} \leq \boldsymbol{x}_{it}; \boldsymbol{\theta}'\boldsymbol{\iota} = 1; \boldsymbol{\theta}, \boldsymbol{z} \geq 0\}. \tag{17.15}$$

The solution to this problem will be the ratio of observed revenue to the maximum revenue possible holding the input vector fixed (i.e., the common measure of revenue efficiency, RE_{it}). Thus, in the special case of the Paasche TFP index, the measure of output-oriented mix efficiency is the well-known measure of revenue-allocative efficiency: $OME_{it} = RAE_{it} = RE_{it}/OTE_{it}$.

The TFP index proposed in this article is the Lowe index given by (17.5), underpinned by the output aggregator function $Q(\boldsymbol{q}_{it}) \propto \boldsymbol{p}_0'\boldsymbol{q}_{it}$. Estimating the output mix-efficiency change component of this TFP index involves solving

$$Q_{it}/\hat{Q}_{it} = \min_{\boldsymbol{\theta},\boldsymbol{z}}\{\boldsymbol{p}_0'\boldsymbol{q}_{it}/\boldsymbol{p}_0'\boldsymbol{z} : \boldsymbol{z} \leq \boldsymbol{Q\theta}; \boldsymbol{X\theta} \leq \boldsymbol{x}_{it}; \boldsymbol{\theta}'\boldsymbol{\iota} = 1; \boldsymbol{\theta}, \boldsymbol{z} \geq 0\}. \tag{17.16}$$

Similar problems are available for identifying the output-oriented mix efficiency levels associated with Laspeyres indexes, and for computing DEA estimates of input-oriented technical efficiency (*ITE*), input-oriented mix efficiency (*IME*), cost efficiency (*CE*) and cost-allocative efficiency (*CAE* = *CE*/*ITE*) associated with Paasche, Laspeyres and Lowe indexes. The efficiency components of the Fisher TFP index can be computed as the geometric average of the Laspeyres and Paasche measures.

17.5 Estimating Maximum TFP and the Rate of Technical Change

The maximum TFP possible using the production technology is the TFP at point E in Fig. 17.1: $TFP_t^* = Q_t^*/X_t^* =$ slope OE. Points of maximum TFP can be estimated using a DEA problem that is closely related to problem (17.14).

Problem (17.14) allows outputs to vary freely but holds the input vector fixed. The point of maximum TFP can be estimated by solving a less restrictive problem that allows both inputs and outputs to vary freely:

$$TFP_t^* = \max_{\boldsymbol{\theta}, \boldsymbol{z}, \boldsymbol{v}} \{Q(\boldsymbol{z}) : \boldsymbol{z} \leq \boldsymbol{Q\theta}; \boldsymbol{X\theta} \leq \boldsymbol{v}; X(\boldsymbol{v}) = 1; \boldsymbol{\theta}' \boldsymbol{\iota} = 1; \boldsymbol{\theta}, \boldsymbol{z}, \boldsymbol{v} \geq 0\} \quad (17.17)$$

where the constraint $X(\boldsymbol{v}) = 1$ is a normalizing constraint that plays the same role as a normalizing constraint in the primal form of (17.12). Again, the solution to problem (17.17) depends on the choices of input and output aggregator functions. Again, if interest centres on identifying the technical change component of a multiplicatively-complete TFP index then the aggregator functions should be those that underpin that index. For example, if the TFP index is the Paasche index then the problem (17.17) becomes

$$TFP_t^* = \max_{\boldsymbol{\theta}, \boldsymbol{z}, \boldsymbol{v}} \{\boldsymbol{p}_{it}' \boldsymbol{z} : \boldsymbol{z} \leq \boldsymbol{Q\theta}; \boldsymbol{X\theta} \leq \boldsymbol{v}; \boldsymbol{w}_{it}' \boldsymbol{v} = 1; \boldsymbol{\theta}' \boldsymbol{\iota} = 1; \boldsymbol{\theta}, \boldsymbol{z}, \boldsymbol{v} \geq 0\}. \quad (17.18)$$

In this case, the optimized value of the objective function is the maximum profitability that can be achieved by firm i in period t. This level of maximum profitability is usually of considerable economic interest, but not for purposes of measuring technical change—as a measure of technical change it is implausible because it varies with observed prices, even when the production possibilities set (represented by the constraints $\boldsymbol{z} \leq \boldsymbol{Q\theta}$ and $\boldsymbol{X\theta} \leq \boldsymbol{v}$) remains unchanged (i.e., even when there is no technical change). In contrast, the technical change component of the Lowe TFP index is robust to differences in the prices faced by different firms. If the TFP index is a Lowe index then problem (17.17) becomes

$$TFP_t^* = \max_{\boldsymbol{\theta}, \boldsymbol{z}, \boldsymbol{v}} \{\boldsymbol{p}_0' \boldsymbol{z} : \boldsymbol{z} \leq \boldsymbol{Q\theta}; \boldsymbol{X\theta} \leq \boldsymbol{v}; \boldsymbol{w}_0' \boldsymbol{v} = 1; \boldsymbol{\theta}' \boldsymbol{\iota} = 1; \boldsymbol{\theta}, \boldsymbol{z}, \boldsymbol{v} \geq 0\}. \quad (17.19)$$

Variations in levels of maximum TFP are due to inward and outward movements in the production frontier in the region of local constant returns to scale (point E in Figs. 17.1 and 17.5). A common view is that these movements are due to variations in technical know-how. However, Solow (1957, p. 312) takes a broader view of technical change and attributes movements in the production frontier to variations in any factors that are not accounted for by the input and output variables that have been included in the analysis. Aside from stocks of scientific and technical knowledge, these "environmental" factors may include anything from measures of labor quality to seasonal conditions. If such environmental factors are favourable for agricultural production then the maximum output possible using any given level of included inputs is higher than the maximum output possible when environmental conditions are poor (i.e., favourable environmental conditions are a form of "technical progress").

Table 17.1 USDA ERS farm production regions

Region	States[a]	Window
1. Pacific	CA, OR, WA	5
2. Mountain	AZ, CO, ID, MT, NM, NV, UT, WY	2
3. Northern Plains	KS, ND, NE, SD	4
4. Southern Plains	OK, TX	8
5. Corn Belt	IA, IL, IN, MO, OH	3
6. Southeast	AL, FL, GA, SC	4
7. Northeast	CT, DE, MA, MD, ME, NH, NJ, NY, PA, RI, VT	2
8. Lake States	MI, MN, WI	5
9. Appalacian	KY, NC, TN, VA, WV	3
10. Delta States	AR, LA, MS	5

[a] Official US Postal Service abbreviations

To account for spatial variations in environmental factors, this article estimates separate variable returns to scale production technologies for the ten farm production regions identified by the USDA-ERS. A list of the states in each region is provided in Table 17.1. To account for temporal variations in environmental factors, these technologies have been estimated using DEA models that allow for a small amount of technical regress. This involves using a moving window of observations to estimate the technology in each region. For example, the Pacific region production technology has been estimated using a moving five-year window of observations, while the Mountain region technology has been estimated using a two-year window. The size of the window was governed by the number of states in each region and reflects a desire to estimate each regional frontier using at least twice as many observations as there are input and output variables in the dataset. The final column in Table 17.1 is the size of the window used to estimate the technology in each region.

Indexes that compare estimated levels of maximum TFP in selected regions are depicted in Fig. 17.8 (Southeast $1960 = 1$). The Southeast region had the highest estimated maximum TFP of any region in both 1960 and 2004 (i.e., $TFP = TFP^* = 1.045$ in Alabama in 1960 and $TFP = TFP^* = 1.896$ in Alabama in 2004). Observe from Fig. 17.8 that the maximum possible TFP in the Southeast region is estimated to have increased by 81.4 % over the sample period (i.e., $\Delta TFP^* = 1.896/1.045 = 1.814$) whereas the maximum TFP in the Southern Plains region is estimated to have increased by only 40 % [i.e., $\Delta TFP^* = (0.946/1.045)/(0.676/1.045) = 0.905/0.646 = 1.40$]. The estimated rate of technical change in the Southern Plains is lower than in the Southeast partly because large windows have the effect of dampening estimated rates of technical change. Indeed, as the size of the window approaches the time-series dimension of the dataset the estimated rate of technical change will approach zero and any TFP change will be attributed entirely to efficiency change.

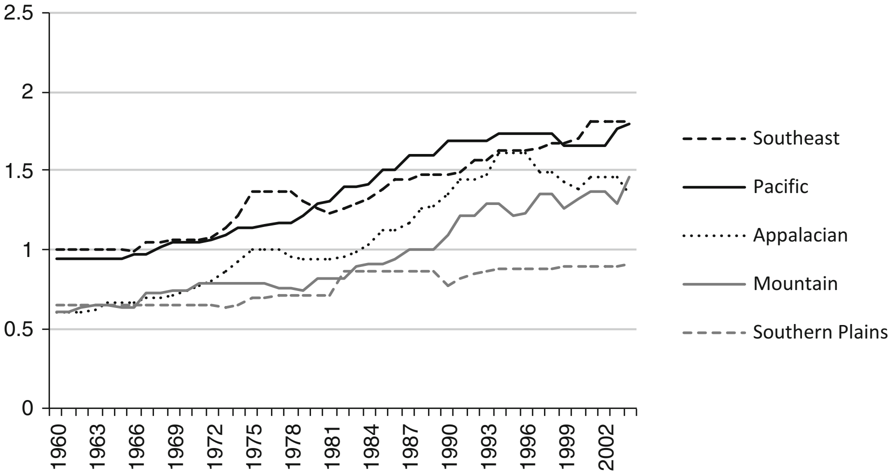


Fig. 17.8 Technical/environmental change in selected regions (Southeast 1960 = 1)

17.6 Data

This article uses a state-level panel dataset compiled by the ERS and recently analysed by, for example, Ball et al. (2004) and Fousekis (2007). Parts of the dataset have also been used recently by LaFrance et al. (2011). The version of the dataset used in this article extends from 1960 to 2004 and comprises observations on the prices and quantities of three outputs (livestock, crops and other outputs) and four inputs (capital, land, labor and materials) in each state in each year. Output quantities are measures of quantities sold plus on-farm consumption and net changes in inventories. Input quantities are measures of purchased inputs as well as farm production used on the farm. Output and input prices are adjusted for subsidies, taxes and direct payments under government commodity programs. Thus, they reflect the net values of outputs and inputs to farmers. The main features of the dataset are described in Ball et al. (1997).

The dataset has its shortcomings. Most importantly, the prices are multilateral price indexes constructed using binary Fisher price indexes and the EKS geometric averaging procedure (Ball et al., 2004). Thus, the price indexes are transitive but fail the identity axiom—the output and input price variables will almost certainly indicate price differences across states and/or time periods even if prices are identical. The quantities are implicit quantity indexes constructed by dividing revenues and costs by the EKS price indexes. This means they satisfy a product test. However, if the price variables do not accurately reflect temporal and spatial variations in prices (because they are intransitive) then it is unlikely that the implicit quantity variables will accurately reflect variations in quantities. Similar problems are found in other U.S. agricultural datasets. For example, the variables in the

InSTePP datasets used by Acquaye et al. (2003) and Alston et al. (2010) are constructed using binary Fisher indexes that violate the transitivity axiom.[6]

17.7 The Components of Profitability and TFP Change in US Agriculture

This section reports selected[7] estimates of changes in agricultural profitability, TFP and efficiency in U.S. agriculture over the 45 years from 1960 to 2004. It has already been reported that in Alabama during this period (1) estimated profitability increased by 4.6 % due to the combined effects of a $1 - 0.504 = 49.6\,\%$ fall in the TT and a 107.6 % increase in TFP, (2) in turn, estimated TFP increased due to an 81.4 % increase in the maximum possible TFP (i.e., "technical change") and a 14.48 % increase in overall efficiency (from 0.549 to 0.629), and (3) in turn, estimated overall efficiency increased due to a 3.2 % increase in output-oriented technical efficiency (from 0.969 to 1) and a 10.9 % increase in output-oriented scale-mix efficiency (from 0.567 to 0.629). These estimates and others (all rounded to two decimal places) are reported in the first rows of Tables 17.2 and 17.3. The remaining rows in these tables report estimates for selected other states. The values that are marked with an "a" are the highest among the 48 states, while those marked with a "b" are the lowest among the 48 states.

The estimates reported in Tables 17.2 and 17.3 are transitive and can therefore be used to make meaningful comparisons of performance across both space and time. For example, the estimates reported in the first column of Table 17.2 reveal that in 1960 Florida was the most profitable state, New Hampshire was the least profitable, and the level of profitability in Alabama was 58 % higher than in New Hampshire ($\Delta PROF = 1.18/0.75 = 1.58$). The second column reveals that by 2004 California had become the most profitable state ($PROF = 1.49$) and West Virginia had become the least profitable ($PROF = 0.55$). The remaining columns reveal that Georgia experienced the largest increase in profitability over the sample period (26 %) on the back of a 157 % increase in TFP, while Tennessee experienced the largest fall in profitability ($-32\,\%$) largely as a consequence of a 60 % deterioration in the TT. Observe that California was the most productive state in 2004 ($TFP = 1.88$) and Wyoming was the least productive ($TFP = 0.52$). The last few columns reveal that Kansas was fully efficient in 1960 and California and Texas were fully efficient in 2004.

[6] Version 4 of the InSTePP dataset covers the period 1949–2002 and can be downloaded from http://www.instepp.umn.edu/data/instepp-USAgProdAcct.html. All variables in this dataset take the value 100 in 1949, so it cannot be used to generate TFP indexes that are comparable with the indexes depicted in Fig. 17.3.

[7] Estimates of profitability change, TFP change, technical change, output-oriented technical efficiency change and output-oriented scale-mix efficiency change in each state in each period are available in a supplementary appendix online.

Table 17.2 Profitability, TFP and efficiency change

	PROF = TT × TFP			TT			TFP			TFP*			TFPE = TFP/TFP*		
State	1960	2004	Δ	1960	2004	Δ	1960	2004	Δ	1960	2004	Δ	1960	2004	Δ
AL	1.18	1.24	1.05	2.06	1.04	0.50	0.57	1.19	2.08	1.05a	1.90a	1.81	0.55	0.63	1.14
CA	1.39	1.49a	1.07	1.58	0.79	0.50	0.88	1.88a	2.14	0.99	1.88	1.91	0.89	1.00a	1.12
FL	1.72a	1.35	0.79	1.91	0.78	0.41	0.90a	1.72	1.92	1.05a	1.90a	1.81	0.86	0.91	1.06
GA	1.02	1.29	1.26a	1.73	0.85	0.49	0.59	1.51	2.57	1.05a	1.90a	1.81	0.56	0.80	1.42
KS	1.19	0.89	0.75	1.92	0.82	0.43	0.62	1.08	1.75	0.62b	1.22	1.98	1.00a	0.88	0.88
LA	1.08	0.82	0.76	2.47	0.81	0.33b	0.44	1.02	2.34	0.66	1.53	2.31	0.66	1.02	
MI	0.82	0.94	1.15	1.82	0.80	0.44	0.45	1.18	2.62	0.68	1.26	1.85	0.66	0.94	1.42
NH	0.75b	0.77	1.03	1.76	0.72	0.41	0.42	1.06	2.51	0.70	1.61	2.30	0.61	1.09	
NM	1.13	1.17	1.03	2.97a	1.36	0.46	0.38	0.86	2.27	0.64	1.53	2.40a	0.60	0.56	0.94
OH	0.76	0.82	1.08	1.38b	0.66	0.48	0.55	1.24	2.27	0.77	1.63	2.12	0.71	1.07	
OK	1.35	0.94	0.70	2.06	1.11	0.54	0.66	0.85	1.29b	0.68	0.95b	1.40b	0.97	0.89	0.92
OR	0.97	1.16	1.20	2.15	0.82	0.38	0.45	1.41	3.12a	0.99	1.88	1.91	0.46b	0.75	1.63a
RI	0.82	0.87	1.07	1.75	0.61b	0.35	0.47	1.42	3.04	0.70	1.61	2.30	0.67	1.32	
TN	1.01	0.68	0.68b	2.14	0.85	0.40	0.47	0.81	1.71	0.63	1.41	2.22	0.74	0.77	
TX	1.17	0.94	0.81	2.29	1.00	0.44	0.51	0.95	1.85	0.68	0.95b	1.40b	0.76	1.00a	1.32
WV	0.78	0.55b	0.71	2.23	0.84	0.38	0.35b	0.66	1.90	0.63	1.41	2.22	0.55	0.47	0.85
WY	1.02	0.86	0.84	2.78	1.65a	0.59a	0.37	0.52b	1.42	0.64	1.53	2.40a	0.58	0.34b	0.59b

Note: a = highest value among the 48 states; b = lowest value among the 48 states

Table 17.3 Output-oriented components of efficiency change

	TFPE = OTE × OSME			OTE			OSE			OME			OSME		
State	1960	2004	Δ	1960	2004	Δ	1960	2004	Δ	1960	2004	Δ	1960	2004	Δ
AL	0.55	0.63	1.14	0.97	1.00a	1.03	0.98	1.00	1.02	0.80b	0.98	1.22a	0.57	0.63	1.11
CA	0.89	1.00a	1.12	1.00a	1.00a	1.00	1.00a	1.00a	1.00	1.00a	1.00a	1.00	0.89	1.00a	1.12
FL	0.86	0.91	1.06	1.00a	1.00a	1.00	1.00a	1.00a	1.00	1.00a	0.94	0.94	0.86	0.91	1.06
GA	0.56	0.80	1.42	0.89	1.00a	1.12	1.00	1.00a	1.00	0.87	1.00	1.14	0.63	0.80	1.26
KS	1.00a	0.88	0.88	1.00a	1.00	1.00	1.00a	1.00	1.00	1.00a	0.92	0.92	1.00a	0.89	0.89
LA	0.66	0.67	1.02	1.00a	1.00a	1.00	1.00a	0.90	0.90	1.00a	1.00a	1.00	0.66	0.67	1.02
MI	0.66	0.94	1.42	1.00a	1.00a	1.00	1.00a	1.00a	1.00	0.92	1.00a	1.09	0.66	0.94	1.42
NH	0.61	0.66	1.09	0.83b	1.00a	1.21a	0.96	1.00a	1.05	0.94	0.85	0.90	0.73	0.66	0.90
NM	0.60	0.56	0.94	1.00a	1.00a	1.00	1.00a	1.00a	1.00	0.99	0.85	0.86b	0.60	0.56	0.94
OH	0.71	0.76	1.07	1.00a	0.93	0.93	1.00a	0.97	0.97	1.00a	0.93	0.93	0.71	0.82	1.15
OK	0.97	0.89	0.92	1.00a	1.00a	1.00	1.00a	1.00a	1.00	1.00a	1.00a	1.00	0.97	0.89	0.92
OR	0.46b	0.75	1.63	1.00a	1.00a	1.00	0.92	1.00a	1.08	1.00a	1.00a	1.00	0.46b	0.75	1.63a
RI	0.67	0.88	1.32	1.00a	1.00a	1.00	0.89b	1.00a	1.13a	1.00a	1.00a	1.00	0.67	0.88	1.32
TN	0.74	0.57	0.77	0.98	0.99	1.01	1.00	0.99	0.99	0.86	0.78b	0.91	0.76	0.58	0.76
TX	0.76	1.00	1.32	1.00a	1.00a	1.00	1.00a	1.00a	1.00	1.00a	1.00a	1.00	0.76	1.00a	1.32
WV	0.55	0.47	0.85	1.00	1.00a	1.00	1.00	0.87	0.87	0.97	1.00a	1.03	0.55	0.47b	0.85
WY	0.58	0.34b	0.59b	1.00a	0.71b	0.71b	0.94	0.96	1.02	0.86	0.79	0.92	0.58	0.48	0.83

Note: a = highest value among the 48 states; b = lowest value among the 48 states

The output-oriented efficiency estimates reported in Table 17.3 indicate that most states were highly technically efficient throughout the sample period (exceptions include New Hampshire in 1960 and Wyoming in 2004). Output-oriented measures of pure scale efficiency and pure mix efficiency were generally high (exceptions include Rhode Island and Alabama in 1960, and Tennessee in 2004) but output-oriented measures of scale-mix efficiency were generally low (especially Oregon in 1960, and West Virginia and Wyoming in 2004). Of course, low levels of scale and/or mix efficiency are not necessarily associated with low levels of profitability (see Fig. 17.1). For example, in 2004 Alabama was one of the most profitable states ($PROF = 1.24$, rank 7 out of 48) and also one of the least scale-mix efficient ($OSME = 0.63$, rank 43).

A slightly different picture of TFP and efficiency change in U.S. agriculture is presented in Table 17.4. This table reports estimated rates of growth in TFP, maximum TFP and measures of efficiency in selected states for the sub-periods 1960–1970, 1970–1980, 1980–1990 and 1990–2002. These particular sub-periods have been chosen to facilitate comparison with estimates reported by Alston et al. (2010) and Ball et al. (1997). The average annual rate of growth in a variable Z between periods s and t can be calculated as $\Delta \ln Z \equiv \ln(Z_t/Z_s)/(t-s)$. For example, the average annual rate of TFP growth in Alabama in the 1960s is estimated to be $\Delta \ln TFP = \ln(TFP_{1970}/TFP_{1960})/(1970-1960) = \ln(0.713/0.574)/10 = 0.0218$ or 2.18 %, and the average annual rate of technical change is estimated to be $\Delta \ln TFP^* = \ln(TFP^*_{1970}/TFP^*_{1960})/10 = \ln(1.12/1.045)/10 = 0.0069$ or 0.69 %. An important property of the estimated growth rates reported in Table 17.4 is that they are additive. This means, for example, that the estimated average annual rate of TFP growth in Alabama in the 1960s is equal to the sum of the technical change and efficiency change components: $\Delta \ln TFP = \Delta \ln TFP^* + \Delta \ln OTE + \Delta \ln OSME = 0.0069 + 0.0031 + 0.0117 = 0.0218$. It also means that estimated average annual rates of TFP and efficiency growth in US agriculture can be computed as arithmetic averages of the estimated growth rates of the 48 states—these US averages are reported in the row labeled US48 in Table 17.4.

At least three features of Table 17.4 are noteworthy. First, the average annual rate of TFP growth in US agriculture is estimated to have been 2.23 % in the 1960s, 0.56 % in the 1970s, 3.06 % in the 1980s, and 1.01 % from 1990 to 2002. These estimated rates of growth are generally quite different from the Alston et al. (2010) and Ball et al. (1997) estimates reported in the last two rows of Table 17.4. Second, the estimated average annual rate of technical change (i.e., change in maximum possible TFP) in US agriculture was 1.84 % in the 1960s, 1.27 % in the 1970s, 2.3 % in the 1980s, and 1.38 % from 1990 to 2002. These nonparametric estimates are similar to parametric estimates reported elsewhere in the literature [e.g., 1.8 % reported by Ray (1982)]. Associated estimated rates of growth in output-oriented technical efficiency (0.06 %, -0.18 %, 0.22 % and -0.02 %) and output-oriented scale-mix efficiency (0.33 %, -0.53 %, 0.54 % and -0.35 %) are relatively small, indicating that technical change has been the major driver of TFP change in each sub-period.

Table 17.4 Average annual rates of growth in TFP and efficiency (%)

	1960–1970				1970–1980				1980–1990				1990–2002			
State	TFP	TFP*	OTE	OSME	TFP	TFP*	OTE	OSME	TFP	TFP*	OTE	OSME	TFP	TFP*	OTE	OSME
AL	2.18	0.69	0.31	1.17	0.26	1.68	−0.13	−1.29	3.37	1.55	0.13	1.69	0.52	1.70	−0.29	−0.89
CA	1.85	1.07	0.00	0.78	2.47	2.09	0.00	0.38	2.67	2.67	0.00	0.00	−0.16	−0.16b	0.00	0.00
FL	1.24	0.69	0.00	0.55	2.31	1.68	0.00	0.64	1.85	1.55	0.00	0.30	1.10	1.70	−0.05	−0.55
GA	3.51	0.69	1.16a	1.66	0.71	1.68	0.00	−0.97	3.34	1.55	0.00	1.79	1.85	1.70	0.00	0.15
KS	1.51	2.24	0.00	−0.73	−0.50	0.82	−1.30	−0.02	3.33	1.99	1.30	0.03	0.21	1.19	0.00	−0.98
LA	4.87a	1.28	0.00	3.60a	−0.50	2.40	−0.76	−2.14b	3.85	2.09	0.76	1.01	−0.02	1.69	0.00	−1.71
MI	2.47	0.83	0.00	1.64	2.72a	1.38	0.00	1.34a	2.37	1.75	0.00	0.62	1.42	1.72	0.00	−0.30
NH	2.60	3.12a	0.09	−0.61	1.11	0.82	1.62a	−1.33	1.86	1.95	0.18	−0.26	1.80	1.37	0.00	0.43
NM	2.47	2.02	0.00	0.44	0.18	0.96	0.00	−0.77	2.25	2.85	−0.08	−0.52	2.38	1.93a	0.07	0.39
OH	2.16	1.28	0.00	0.88	0.97	0.51b	−0.11	0.58	2.99	2.59	0.11	0.29	0.76	1.81	0.00	−1.05
OK	−1.30b	0.00b	0.00	−1.30	1.63	1.03	0.00	0.61	1.29b	0.85b	0.00	0.43	1.26	1.20	0.00	0.06
OR	3.27	1.07	0.00	2.20	2.49	2.09	−0.24	0.64	2.35	2.67	0.24	−0.56	0.78	−0.16b	0.00	0.94a
RI	3.99	3.12a	0.00	0.87	−0.35	0.82	0.00	−1.18	5.76a	1.95	0.00	3.81a	0.23	1.37	0.00	−1.14
TN	2.00	2.03	0.18	−0.22	0.52	2.42a	−2.16	0.25	3.05	3.56a	1.92	−2.43b	−0.39	0.63	−0.60b	−0.42
TX	1.05	0.00b	0.00	1.05	0.76	1.03	−0.77	0.50	2.70	0.85b	0.77	1.08	1.31	1.20	0.00	0.11
WV	1.54	2.03	0.01	−0.49	1.81	2.42a	0.00	−0.61	3.00	3.56a	0.00	−0.57	−0.21	0.63	0.00	−0.84
WY	1.60	2.02	−1.18b	0.75	0.08	0.96	−0.89	0.02	1.80	2.85	0.85	−1.90	−0.54b	1.93a	−0.30	−2.17
US48	2.23	1.84	0.06	0.33	0.56	1.27	−0.18	−0.53	3.06	2.30	0.22	0.54	1.01	1.38	−0.02	−0.35
A	2.23	1.64	0.03	0.56	0.56	2.47	−0.44	−1.47	3.06	2.67	0.13	0.27	1.01	−0.05	−0.06	1.12
B	2.23	0.00	−0.32	2.55	0.56	0.00	−0.76	1.32	3.06	0.00	2.16	0.89	1.01	0.00	0.76	0.25
Alston	1.69				2.46				2.07				0.97			
Ball	1.91				1.22				3.36							

Note: a = highest value among the 48 states; b = lowest value among the 48 states

Finally, this article has used the term “technical change” to refer to both spatial and temporal variations in the production environment. To accommodate this broad concept of technical change, ten different production frontiers (one for each ERS farm production region) were estimated in each time period. A common alternative estimation approach involves estimating a single frontier in each period—such a model allows for temporal variations in the production environment, but does not allow for spatial variations. Estimates of TFP and efficiency change obtained using this more restrictive “Model A” are summarized in row A in Table 17.4. Unlike some other TFP indexes (e.g., the Malmquist, Hicks-Moorsteen and Diewert-Morrison indexes), the Lowe TFP index can be computed without knowing (or assuming) anything about the production technology. Thus, different restrictions on the nature of technical change will only affect the estimated technical change and efficiency change components of TFP change (i.e., the TFP index itself will remain unaffected). The results from Model A reported in Table 17.4 indicate that prior to 1990 technical progress was the most important driver of TFP growth in U.S. agriculture. However, after 1990 output-oriented scale-efficiency change became the most important driver. An even more restrictive model is a “no technical change” model that involves using all the observations in the dataset to estimate a single frontier. Results obtained using this “Model B” are summarized in row B in Table 17.4. By design, Model B attributes all TFP growth to improvements in various types of efficiency—this models suggests that prior to 1980 the main driver of TFP change was scale-mix efficiency change; after 1980 it was technical efficiency change. A discussion of statistical methods for choosing between different types of models is beyond the scope of this article.

17.8 Conclusion

This article uses an aggregate quantity-price framework to decompose agricultural profitability change into a measure of TFP change and a measure of change in the agricultural TT. It argues that deteriorations in the (expected) TT will generally be associated with improvements in agricultural TFP. To illustrate, the article estimates that, over the period 1960–2004, a 50 % decline in the agricultural TT in Alabama has been associated with a two-fold increase in TFP.

Well-known measures of TFP change include Laspeyres, Paasche, Fisher and Törnqvist TFP indexes. A problem with these indexes is that they fail a commonsense transitivity axiom. The usual way around the problem is to apply a geometric averaging procedure known as the EKS method. Unfortunately, so-called EKS indexes fail an identity axiom. This article proposes a new Lowe TFP index that satisfies both the transitivity and identity axioms. It also demonstrates that the choice of index formula matters—in one instance the EKS index indicates that two states were equally productive but the Lowe index indicates that one state was 39 % more productive than the other.

The Lowe index is a member of the class of multiplicatively-complete TFP indexes. In O'Donnell (2008) I demonstrate that, in theory, all such indexes can be exhaustively decomposed into a measure of technical change and various measures of technical, scale and mix efficiency change. Technical change is a measure of movements in the production frontier, technical efficiency change is a measure of movements towards the frontier, and scale and mix efficiency change are measures of movements around the frontier surface to capture economies of scale and scope. This article shows how the methodology can be used to decompose the Lowe TFP index. It finds that the main driver of TFP change in U.S. agriculture over the sample period has been technical progress (e.g., at an annual average rate of 1.84 % in the 1960s and 2.30 % in the 1990s), that levels of technical efficiency have been stable and high (e.g., the geometric mean of all $it = 2160$ OTE estimates is 0.989), and that levels of scale-mix efficiency have been highly variable and relatively low (e.g., the geometric mean of the OSME estimates is 0.791). These findings support the view that research and development expenditure has led to expansions in the production possibilities set, that U.S. farmers adopt new technologies quickly and make relatively few mistakes in the production process, and that they rationally adjust the scale and scope of their operations in response to changes in prices and other production incentives.

Further research in at least two areas is needed to substantiate these general findings. First, there is a need to re-evaluate the index number methods that are used to construct price and quantity variables at disaggregated levels. The (binary) Fisher and Törnqvist indexes that are currently used to construct agricultural datasets (including the dataset used in this article) are no better suited to constructing (multilateral) indexes of livestock and crop outputs, for example, than they are to constructing (multilateral) indexes of aggregate output or TFP. Second, non-parametric and/or parametric statistical procedures (e.g., bootstrapping, hypothesis tests) should be used to assess the plausibility of different assumptions concerning the production technology and the nature of technical change. These assumptions play a key role in identifying the components of TFP change that policy-makers need. For example, a constant returns to scale assumption is enough to rule out any estimated TFP gains associated with improvements in scale efficiency, estimation of an average production function rather than a frontier function is enough to rule out any estimated TFP gains associated with improvements in technical efficiency, and this article has shown how different assumptions concerning changes in the production environment can be used to partially or totally eliminate the estimated TFP gains associated with technical change.

References

Acquaye A, Alston J, Pardey P (2003) Post-war productivity patterns in U.S. agriculture: influences of aggregation procedures in a state-level analysis. Am J Agric Econ 85:59–80

Alston J, Andersen M, James J, Pardey P (2010) Persistence pays: U.S. agricultural productivity growth and the benefits of public R&D spending. Springer, New York

Ball V, Bureau JC, Nehring R, Somwaru A (1997) Agricultural productivity revisited. Am J Agric Econ 79:1045–1063
Ball V, Hallahan C, Nehring R (2004) Convergence of productivity: an analysis of the catch-up hypothesis within a panel of states. Am J Agric Econ 86:1315–1321
Capalbo S (1988) Measuring the components of aggregate productivity growth in U.S. agriculture. West J Agric Econ 13:53–62
Caves D, Christensen L, Diewert W (1982) The economic theory of index numbers and the measurement of input, output, and productivity. Econometrica 50:1393–1414
Diewert W, Morrison C (1986) Adjusting output and productivity indexes for changes in the terms of trade. Econ J 96:659–679
Elteto O, Koves P (1964) On a problem of index number computation relating to international comparison. Stat Szle 42:507–518
Farrell M (1957) The measurement of productive efficiency. J R Stat Soc Ser A 120:253–290
Fousekis P (2007) Growth determinants, intra-distribution mobility, and convergence of state-level agricultural productivity in the USA. Int Rev Econ 54:129–147
Griliches Z (1961) An appraisal of long-term capital estimates: comment. Working paper, Princeton University Press
Hill P (2008) Lowe indices. In: 2008 world congress on national accounts and economic performance measures for nations, Washington
Jorgenson D, Griliches Z (1967) The explanation of productivity change. Rev Econ Stud 34:249–283
LaFrance J, Pope R, Tack J (2011) Risk response in agriculture. NBER Working Paper Series No. 16716, NBER, Cambridge
Lowe J (1823) The present state of England in regard to agriculture, trade and finance, 2nd edn. Longman, Hurst, Rees, Orme and Brown, London
Morrison Paul C, Nehring R (2005) Product diversification, production systems, and economic performance in U.S. agricultural production. J Econom 126:525–548
Morrison Paul C, Nehring R, Banker D (2004) Productivity, economies, and efficiency in U.S. agriculture: a look at contracts. Am J Agric Econ 86:1308–1314
O'Donnell C (2008) An aggregate quantity-price framework for measuring and decomposing productivity and profitability change. Centre for Efficiency and Productivity Analysis Working Papers No. WP07/2008, University of Queensland. http://www.uq.edu.au/economics/cepa/docs/WP/WP072008.pdf
O'Donnell C (2010) Measuring and decomposing agricultural productivity and profitability change. Aust J Agric Resour Econ 54:527–560
O'Donnell C, Shumway C, Ball V (1999) Input demands and inefficiency in U.S. agriculture. Am J Agric Econ 81:865–880
Ray S (1982) A translog cost function analysis of U.S. agriculture, 1939–77. Am J Agric Econ 64:490–498
Solow R (1957) Technical change and the aggregate production function. Rev Econ Stat 39:312–320
Szulc B (1964) Indices for multi-regional comparisons. Prezeglad Statystyczny (Stat Rev) 3:239–254

Chapter 18
Research Fronts and Prevailing Applications in Data Envelopment Analysis

John S. Liu, Louis Y.Y. Lu, and Wen-Min Lu

Abstract Research activities relating to data envelopment analysis (DEA) have grown at a fast rate recently. Exactly what activities have been carrying the research momentum forward is a question of particular interest to the research community. This study finds these research activities, or research fronts, as well as some facts on applications in DEA. A research front refers to a coherent topic or issue addressed by a group of research articles in recent years. The large amount of DEA literature makes it difficult to use any traditional qualitative methodology to sort out the matter. Thus, this study applies a network clustering method to group the literature through a citation network established from the DEA literature over the period 2000–2014. The keywords of the articles in each discovered group help pinpoint its research focus. The four research fronts identified are "bootstrapping and two-stage analysis", "undesirable factors", "cross-efficiency and ranking", and "network DEA, dynamic DEA, and SBM". Each research front is then examined with key-route main path analysis to uncover the elements in its core. In the end, we present the prevailing DEA applications and the observed association between DEA methodologies and applications.

Keywords Data envelopment analysis • Research front • Literature survey • Main path analysis

J.S. Liu (✉)
Graduate Institute of Technology Management, National Taiwan University of Science and Technology, 43, Section 4, Keelung Road, Taipei 10607, Taiwan
e-mail: johnliu@mail.ntust.edu.tw

L.Y.Y. Lu
College of Management, Yuan Ze University, 135 Yuan-Tung Road, Chung-Li, Taoyuan 32003, Taiwan
e-mail: louislu@saturn.yzu.edu.tw

W.-M. Lu
Department of Financial Management, National Defense University, 70, Section 2, Zhongyang North Road, Beitou, Taipei 112, Taiwan
e-mail: wenmin.lu@gmail.com

J. Zhu (ed.), *Data Envelopment Analysis*, International Series in Operations Research & Management Science 238, DOI 10.1007/978-1-4899-7684-0_18

18.1 Introduction

Data envelopment analysis (DEA) is a methodology for performance evaluation and benchmarking where multiple performance measures are present (Cook et al. 2014; Sherman and Zhu 2013). Charnes et al. (1978) establish an efficient frontier formed by the best performing decision making units (DMUs) and assign an efficiency index to each non-frontier units according to their distances to the efficient frontier. After more than 35 years of development, research activities relating to DEA are still growing at a very fast rate. Our search in the Web of Science (WOS) database indicates that during the period 2010–2014, around 2000 more DEA-related papers have been published, which is in addition to the 4500 existing collections prior to 2010 as reported by Liu et al. (2013a).

Amidst such a large amount of literature, it is getting more and more difficult to comprehend the development of the field without the guidance of survey type studies. Several previous studies have reviewed the general DEA literature. They include, in the sequence of their publication year, Seiford and Thrall (1990), Seiford (1996), Gattoufi et al. (2004a, b), Cooper et al. (2007), Emrouznejad et al. (2008), Cook and Seiford (2009), Liu et al. (2013a, b), etc. Seiford and Thrall (1990) review DEA development when the field was in its early stage. Seiford (1996) and Gattoufi et al. (2004a) extensively list the DEA literature up to the time when their articles was published. Cooper et al. (2007) review DEA models and measures from a theoretical perspective. Gattoufi et al. (2004b) and Emrouznejad et al. (2008) conduct bibliometric style surveys and present the DEA publication statistics. Cook and Seiford (2009) present a comprehensive review on the methodological developments since 1978. Liu et al. (2013a, b) survey the general DEA literature and examine how DEA is applied through a citation-based methodology.

There is also no lack of survey studies in specific DEA subfields. Studies that survey methods to further discriminate DEA results include Adler et al. (2002), Angulo-Meza and Lins (2002), and Hosseinzadeh Lotfi et al. (2013). Zhu (2003) and Hatami-Marbini et al. (2011) review the approaches to handle imprecise and fuzzy data in DEA. Zhou et al. (2008) and Song et al. (2012) survey energy and environmental studies that apply DEA. For financial institution and banking applications, Berger and Humphrey (1997), Fethi and Pasiouras (2010), and Paradi and Zhu (2013) conduct comprehensive surveys. Recently, Cook et al. (2010a), Castelli et al. (2010), and Kao (2014b) go over the developments in network DEA and two-stage process model.

All these surveys enrich the DEA literature and help advance our understanding of the field. They provide guidance to scholars who are new to the field and may have helped widen the view of long standing researchers. This current study attempts to play the same role as those previous surveys by applying a quantitative and systematic approach similar to that adopted in Liu et al. (2013a). In fact, this current study continues and complements the previous survey by Liu et al. (2013a) with up-to-date data. As regards to methodology, in addition to main path analysis, this study also adopts a widely used network clustering method to separate the literature into groups. Nevertheless, we deviate from Liu et al. (2013a) in that this study's main focus is on research fronts of DEA. Aside from providing updates on the main trajectory of the overall DEA development and authors who have made significant contributions to the development, the current study answers an interesting question: What are the recent major activities that have carried the DEA research momentum forward?

Research fronts are areas in which many researchers put their focus in recent years. They are typically represented by a group of articles addressing the same or similar issues. In order to discover these coherent groups within the DEA field, we construct a citation network and apply a powerful network clustering method, the edge-betweenness based clustering (Girvan and Newman 2002), to separate the citation network into sub-networks. The clustering method demands that the resulting sub-networks have nodes tightly knitted within, but loosely connected to nodes outside the sub-network. We then apply key-route main path analysis assisted by keyword analysis to comprehend the research activities in each group. As to the time coverage, Liu et al. (2013a) observes two phases of DEA development with the latest phase beginning around 2000. It is quite reasonable to take that year as the starting point, and thus we take articles from 2000 through to 2014 to uncover research fronts.

The growth of DEA research is largely boosted by its broad applicability. Scholars have applied various DEA methodologies to study efficiencies and related management issues in a wide array of applications. Applications in different industry, nevertheless, seemed to be associated with certain methodologies. Some of such phenomena are rooted in the nature of applications while others are not. For those that are not, comprehending the phenomena opens opportunities for further innovation in the application research. The last part of this study applies keyword analysis to find association between methodologies and applications, along with some basic descriptive analysis on applications.

This paper is organized as follows. After the introduction, we briefly describe the methodology used in this study, in particular the edge-betweenness based clustering and the key-route main path analysis. Section 18.3 discusses data composition and presents the basic statistics. Section 18.4 offers the results of overall DEA development, which is essentially an update of the major results in (Liu et al. 2013a). Section 18.5 elaborates on the DEA research fronts. It is followed by a presentation of keyword analysis results on DEA applications in particular the association between methodologies and applications. The last section concludes and discusses future research avenues to pursue.

18.2 Methodologies

The methodologies to uncover research fronts are based on citation relationships among articles, or more exactly a citation network. In a citation network, each node represents an article, and it is linked to other nodes that it references or is cited by. Using network terminology, a citation network is a non-weighted and directed network. It is non-weighted, because the importance of each citation is regarded as the same. It is directed, as presumed knowledge flows directionally from a cited article to the article that references it.

This study adopts two methods to analyze the citation network established from the collected dataset. We apply the first method, the edge-betweenness based clustering (Girvan and Newman 2002) in association with the optimal modularity concept (Newman 2006), to find coherent groups in the citation network. The second method, key-route main path analysis (Liu and Lu 2012), helps examine the content of each group. The following sections briefly introduce these two methods.

18.2.1 Edge-Betweenness Based Clustering

Grouping articles that address similar issues is an essential step in discovering research fronts. We assume that citation relationships among articles can be used to determine any similarity between articles. If two articles reference several of the same articles and are also cited by several of the same other articles, then the probability of these two articles addressing the same issue is high; otherwise, they probably study quite different problems. Thus, articles that address the same issues form a tightly knitted 'community' in a citation network. Based on such a premise, this study adopts the widely used edge-betweenness based clustering (Girvan and Newman 2002) with the assistance of the optimal modularity concepts (Newman 2006) to group similar articles. Both the edge-betweenness clustering and modularity concepts are originated from physicist Mark Newman.

The modularity for a network is defined as "the number of edges (links) falling within groups minus the expected number in an equivalent network with edges placed at random" (Newman 2006). A network with high modularity is dense in connections between the nodes within groups, but sparse in connections between nodes in different groups. The best division for a network is the one that has the largest value of network modularity.

The edge-betweenness of a network link is "the number of shortest paths between pairs of vertices that run along it" (Girvan and Newman 2002). The links that connect groups in the network have high edge-betweenness. The edge-betweenness based clustering relies on this property and removes step by step links with the highest edge-betweenness. In the process, the groups are separated from one another and the underlying group structure is gradually revealed.

The clustering procedure begins with a calculation of edge-betweenness of all existing links in the network. The link with the highest edge-betweenness is then removed. One then recalculates edge-betweenness for the links that are affected by the previous removal, which is followed by removal of the current highest-rank link. Eventually, the network is divided into two groups. At this step, the modularity for such division is computed and recorded. The recalculation and removal steps continue. The modularity is calculated whenever there are existing groups that are divided further. The recalculation and removal steps are repeated until all the links in the network are removed. At this point, one traces back to the network division that has the largest modularity and obtains the grouping result. In practice, we adopt igraph (Csardi and Nepusz 2006) implementation of the algorithm under Microsoft Visual Studio development environment.

18.2.2 Key-Route Main Path Analysis

One of the more widely used extensions of main path analysis is the key-route main path approach (Liu and Lu 2012), which extends the main path analysis originally introduced by Hummon and Dereian (1989). A main path is a connected chain of significant links in an acyclic directed network such as a citation network. In general, the procedure for main path analysis consists of two steps. First, it finds the significance measure "traversal count" for each network link based on the network structure. Search path link count (SPLC) is the algorithm adopted in this study to measure the significance of the links. Second, it searches for the main path(s) based on traversal counts found in the first step. Here, the method as suggested by Hummon and Dereian (1989) searches only forward—that is, it establishes paths from a given network node by moving along the direction of links. The path(s) found this way run the risk of missing many top significant links. The key-route main path approach proposed by Liu and Lu (2012) begins the search from the top significant links, thus guaranteeing the inclusion of these links. It requires specifying the number of the top links (key-routes). For example, key-route 10 specifies that the search begins from the top 10 most significant links. By varying the number of key-routes, one is able to control the level of main path details.

The search in key-route main path analysis starts by picking one top link and searching backward from the tail node of a given link as well as forward from the head node of the same link. The search can be global or local (Liu and Lu 2012). This study adopts the global search, which means connecting the link that deliver the largest overall traversal count value. A path is obtained by joining both forward and backward search results with the link itself. One repeats the same process for all the specified top links. The final key-route main paths are a union of all the paths thus obtained. The key-route main path analysis has been successfully applied to detect technological changes (Chen et al. 2013a; Hung et al. 2014), conducting literature reviews (Chuang et al. 2014; Lu and Liu 2013, 2014), and tracing the decisions of legal opinions (Liu et al. 2014), etc.

18.3 Data

This study obtains data from the Thomson Reuters Web of Science (WOS) database service. Data are separated into two parts in time. Part I ranges from 1978 to 1999, whereas Part II ranges from 2000 to 2014. We use the combination of Parts I and II data to observe the overall DEA development and Part II data to examine the research front. Part I data were taken from the dataset used in a previous study by Liu et al. (2013a) and retrieved on August 2010. Part II data were retrieved from WOS on July 17, 2014.[1] Since research front is the main objective of this study and it is more likely to be observed from articles published in prestigious journals and highly cited articles, we limit our data to articles published in several prestigious journals as well as top cited articles. In other words, this study considers only articles that are either published in prestigious journals or are highly cited in the WOS database. Certainly, many articles meet both criteria.

In our preliminary dataset, there are more than 400 journals that published at least two DEA-related articles in the period 2000–2014. To select journals, we consulted two journal rating sources[2] and conducted a survey with experts in the field. Twenty experts were issued a survey letter and ten of them responded. In the end, 12 journals are selected, which are listed in Table 18.1.[3] All the journals selected receive good rankings in the two journal rating sources. More importantly,

Table 18.1 Target journals (in alphabetical order)

	Journals	Number of expert endorsements[a]
1	Annals of Operations Research	9
2	Applied Economics	7
3	Computers & Operations Research	6
4	European Journal of Operational Research	10
5	International Journal of Production Economics	5
6	Journal of Banking & Finance	7
7	Journal of Econometrics	5
8	Journal of Productivity Analysis	10
9	Journal of the Operational Research Society	10
10	Management Science	8
11	Omega-International Journal of Management Science	10
12	Operations Research	9

[a]A total of 10 experts responded to the survey

[1] Part II data thus contain only articles in the first half of 2014.

[2] The two sources are the Harzing Journal Quality List (http://www.harzing.com/pop.htm) and The ABDC Journal Quality List (http://www.abdc.edu.au/pages/abdc-journal-quality-list-2013.html).

[3] In the same survey, several experts indicate that they do not feel comfortable with DEA papers published in some journals. Among them, African Journal of Business Management, Applied Mathematics and Computation, and Expert Systems with Applications are high on the list.

they are all endorsed by more than five of the experts we consulted as journals that publish quality DEA articles.

In addition to articles published in the 12 selected journals, this study includes the top 400 cited articles in Part I data and the top 500 articles in Part II data. In practice, the data search procedure first filters-in the top cited articles and then takes in those that are published in the selected journals from the remaining articles.[4] In the end, Part I data consist of 733 articles and Part II data consist of 1595 articles. The sum of both, 2328, is used to analyze overall DEA development.

18.4 Overall DEA Development

Liu et al. (2013a) conduct a citation-based literature survey using data collected in August 2010. The DEA literature has grown tremendously since then. As mentioned in the introduction section, around 2000 DEA papers have been published in the period 2010–2014. This section provides an update to the major results of Liu et al. (2013a). It should be kept in mind, however, that there is a difference in data criteria between this study and the previous one. The previous study includes all DEA articles in the analysis, whereas this study takes only the top cited articles and those articles published in the 12 selected journals.

18.4.1 Researcher Statistics

This section presents top DEA researchers in order to recognize their significant contributions to the field. In the WOS database, some authors are listed by different identities, e.g., by showing or not showing their middle initials. Before identifying the top researchers, we correct authors' names so that each author has only one identity.

Table 18.2 lists the top 29 DEA authors in the order according to their g-index (Egghe 2006) followed by h-index (Hirsch 2005). It can be regarded as an update to Table 18.1 of Liu et al. (2013a), which covers researchers in the period 1978–2010, while the current table encompasses scholars in the period 1978–2014. The indices are calculated based on citation number listed in the WOS database on April 21, 2015. We also expand the list from top 20 to top 29 in order to maintain the same cutoff point, which is at g-index 19. The list contains certainly DEA pioneers Charnes, Cooper, Banker, etc. One notices that the indices for Charnes are smaller than those in the previous study (Liu et al. 2013a). This is because this current study

[4] The procedure preserves medium to high cited articles even though they are not in the 12 selected journals.

Table 18.2 Top 29 DEA researchers (1978–2014) according to their *g*-index

	Authors	*g*-index	*h*-index	Years active	Total number of papers
1	Cooper, William W	71	35	1978–2014	71
2	Zhu, Joe	49	30	1995–2014	71
3	Färe, Rolf	48	23	1978–2014	48
4	Grosskopf, Shawna	44	26	1983–2013	44
5	Cook, Wade D.	43	25	1985–2014	55
6	Banker, Rajiv D	42	26	1980–2010	42
7	Thanassoulis, Emmanuel	40	21	1987–2014	41
8	Sueyoshi, Toshiyuki	38	30	1986–2013	51
9	Charnes, A	35	26	1978–1997	35
10	Seiford, Lawrence M.	34	26	1982–2009	34
11	Lovell, C.A. Knox	31	20	1978–2012	31
12	Kao, Chiang	31	18	1991–2014	36
13	Simar, Leopold	28	20	1995–2014	28
14	Pastor, Jesus T.	26	14	1995–2014	26
15	Liang, Liang	25	15	2006–2014	25
16	Tone, Kaoru	25	14	1996–2014	25
17	Thrall, R.M.	24	15	1986–2004	24
18	Golany, Boaz	23	19	1985–2008	23
19	Chen, Yao	22	15	2002–2013	28
20	Podinovski, Victor V.	22	12	1997–2013	25
21	Wilson, Paul W.	21	16	1993–2012	21
22	Forsund, Finn R.	21	13	1979–2014	21
23	Kuosmanen, Timo	21	13	2000–2014	32
24	Paradi, Joseph C.	21	13	1997–2014	21
25	Ruggiero, John	20	11	1996–2014	25
26	Dyson, Robert G.	19	16	1987–2010	19
27	Athanassopoulos, A.D.	19	14	1995–2004	19
28	Ray, Subhash C.	19	10	1988–2014	19
29	Camanho, Ana S.	19	9	1999–2013	19

includes fewer articles in general for all authors and that Charnes did not produce any more DEA papers after 1997. In comparison with the list in the previous study, the table now includes several new names.

18.4.2 Main Paths

Figure 18.1 presents the key-route main paths of DEA development from 1978 to 2014. The number of top key-routes is set to 20. In the figure, the arrow indicates the direction of knowledge flow, and the line thickness reflects the size of its

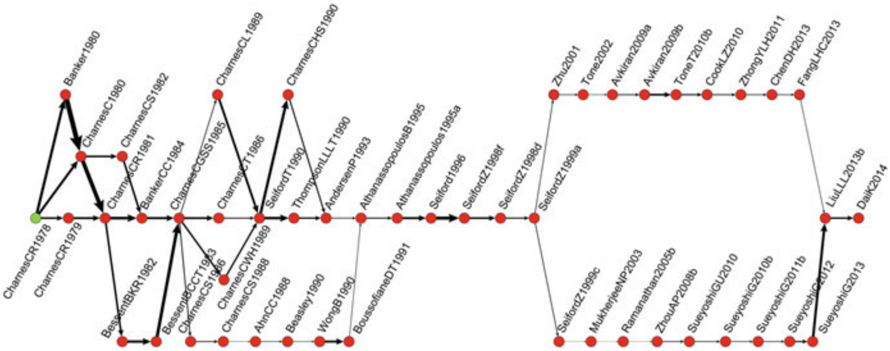


Fig. 18.1 Key-route main path of overall DEA development (for top 20 key-routes). Link weights are indicated with different line thickness. *Thicker lines* indicate heavier weights

traversal count. The thicker the line is, the more significant the route is. The figure is drawn with Pajek software (Batagelj and Mrvar 1998). The "key-route 20" main paths consist of 50 papers. Each paper in the figure is attached with a notation that begins with the last name of the 1st author followed by the 1st letters of the co-author's last name and ends with the publication year of the paper. For example, the original paper by Charnes, Cooper and Rhodes in 1978 is indicated as CharnesCR1978.

The main paths take the shape of a tennis racket with one central path (handle) followed by a hoop that consists of papers in various subareas of DEA. There is no surprise that the main path begins from CharnesCR1978 (Charnes et al. 1978), which introduces the renowned CCR model. It then passes through the article that proposes the BCC model (BankersCC1984 (Banker et al. 1984)) and continues to SeifordZ1999a (Seiford and Zhu 1999a) before diverging into two paths. One can see that both the local and global main paths reported in Liu et al. (2013a) are embedded in the paths between CharnesCR1978 to Seiford1996 (Seiford 1996). The similarity between the previous and the current results ends at Seiford1996. New research activities after 2010 change the main paths. Following Seiford1996, which is a review article, SeifordZ1998f (Seiford and Zhu 1998b), SeifordZ1998d (Seiford and Zhu 1998a), and SeifordZ1999a make up a series of studies on sensitivity analysis and super-efficiency models.

Following SeifordZ1999a are two paths, each one including papers that focus on various topics in DEA. In a scientific field with a dominant subarea, all papers on the main paths can belong to the same subarea. On the other hand, in a scientific field with several subareas of roughly equal awareness, as is the case we have seen here for DEA, one can have papers in different subareas alternating on the main path. The subject of interest of the papers on the paths actually hints that the subareas are significant in the historical development of DEA.

The two paths following SeifordZ1999a contain papers discussing the subject of super-efficiency, slacks-based measure, network DEA, dynamic DEA (upper path), as well as banking and environmental studies (lower path). On the upper path, Zhu2001 (Zhu 2001) continues the development of applying a super-efficiency concept for the sensitivity analysis. Tone2002 (Tone 2002) introduces a slacks-based measure of super-efficiency. Avkiran2009a (Avkiran 2009b) proposes a four-stage approach designed to remove the impact of exogenous factors. Avkiran2009b (Avkiran 2009a) demonstrates network DEA in a slacks-based measure format using UAE bank data. ToneT2010b (Tone and Tsutsui 2010) develops a dynamic DEA model in a slacks-based measure framework. CookLZ2010 (Cook et al. 2010a) reviews the two-stage process DEA models that are a special case of network DEA models. ZhongYLH2011 (Zhong et al. 2011) investigates R&D performance in China. ChenDH2013 (Chen et al. 2013c) and FangLHC2013 (Fang et al. 2013) present novel variations on the super-efficiency model.

The lower path consists of two articles on banking and a series of articles that study environmental performance. SeifordZ1999c (Seiford and Zhu 1999b) is the first to introduce a two-stage process concept to study bank performance. MukherjeeNP2003 (Mukherjee et al. 2003) applies a similar two-stage process concept to measure the efficiency of banking service in India. The topic then turns to methods to deal with undesirable factors which are essential common in energy and environmental studies. Ramanathan2005b (Ramanathan 2005) applies DEA to forecast energy consumption and CO2 emissions. ZhouAP2008b (Zhou et al. 2008) conducts a comprehensive literature survey on the application of DEA to energy and environmental studies. It is followed by a series of works by Sueyoshi and colleagues (SueyoshiGU2010 (Sueyoshi et al. 2010), SueyoshiG2010b (Sueyoshi and Goto 2010b), SueyoshiG2011b (Sueyoshi and Goto 2011a), SueyoshiG2012 (Sueyoshi and Goto 2012), and SueyoshiG2013 (Sueyoshi and Goto 2013)), which evaluate performances of coal-fired power plants, petroleum industry, etc. with an emphasis on the methods used to deal with undesirable (bad) outputs.

The two paths cascade to LiuLLL2013b (Liu et al. 2013b), which surveys DEA applications through a bibliometric method. DaiK2014 (Dai and Kuosmanen 2014) proposes an approach that combines DEA with clustering methods to benchmarking DMUs.

The key-route main paths highlight many highly recognized works in DEA development. No doubt, many important works are not revealed, but it does sketch the outlines of a grand DEA development especially in the early stage. The key-route main path approach expands the traditional method, thus providing us a richer view of the development trajectory. After SeifordZ1999a, the main paths include discussions on various subjects of DEA. Are these discussions truly representative of recent DEA research activities? The next section investigates further to answer the question.

18.5 DEA Research Fronts

The key-route main paths presented in the previous section highlight certain aspects of overall DEA development, but may not reveal enough detail of the research fronts. This section focuses on more recent articles and applies the edge-betweenness clustering method to Part II dataset (2000–2014) in order to discover active DEA research subareas in recent years. One issue needs to be addressed before clustering, and that is whether to include in the analysis survey type articles in which their discussions encompass the whole DEA field. Survey type articles impose a difficulty in clustering, because they logically do not belong to any of the subareas. There are four such articles (Cook and Seiford 2009; Cooper et al. 2007; Liu et al. 2013a, b) during the period 2000–2014. In order to know the difference these four articles can make, we conduct a pilot analysis that analyzes two datasets. One includes all articles in Part II dataset while the other has the four survey articles removed. The results show that the network modularity of the two analyses is 0.513 and 0.525, respectively, out of scale from 0 to 1. The one without the four review articles provides a better clustering result, as indicated by its higher network modularity, and thus we adopt it in this study.

Edge-betweenness clustering divides the DEA citation network into groups of various sizes. The largest group contains 156 articles, while the smallest includes only one article. Four groups have a size greater than 100, consisting of 156, 156, 152, and 147 articles, respectively. The total number of papers in the top four groups is 611, which amounts to 38.3 % of Part II data. The sizes of the groups ranked number 5th to 10th are 97, 55, 47, 47, 38, and 32. Taken together, the top 10 groups contain 927 (58.1 %) of Part II data. The remaining groups are of size 29 and smaller. Among them, 91 groups are of size less than 5. The majority of these small-size groups are either isolated nodes or ‘islands’ in the citation network. Reporting results with so many small size groups is actually one of the advantages of edge-betweenness clustering methods—it does not enforce an attachment of an entity to a seemingly irrelevant group, like what is done in the K-means method. It leaves remotely relevant entities alone as small groups.

Those groups of a larger size are the subareas that have a large amount of coherent research activities and can therefore be regarded as the research fronts in DEA. In the following discussions, we concentrate on the four largest groups and apply key-route main path analysis to each group in order to identify the essence of each research front. The key-route main paths are created from the top 10 -key-routes. The groups ranked number 5th–10th, however, deserve brief notes. They are discussed together at the end of this section.

To comprehend the contents of each group, we count keywords in the titles and abstracts of the articles in each group. The keywords are selected from a list of terms used in all the articles excluding stop-words[5] and words that are too common

[5] Stop words include ‘about’, ‘become’, ‘could’, ‘give’, ‘please’, ‘seems’, ‘under’, ‘whereas’, etc., to name a few.

in DEA such as ‘DEA’, ‘efficiency’, ‘model’, ‘measure’, etc. Variations of terms with identical or similar meanings are regarded as the same keywords—for example, ‘cross efficiencies’, ‘cross efficiency’, ‘cross-efficiencies’, and ‘cross-efficiency’ are all treated as the same keywords.

Table 18.3 presents the results of the keyword analysis. Keywords within each group are listed in the order of their appearance count. Only keywords with an average appearance count greater than 0.12 are listed. Except for the 1st and 4th groups, the focus for each group is identified from the keywords without much struggle. For example, the 2nd group includes mostly terms related to energy and environment and undesirable factors, while ‘cross-efficiency’ is the prevailing term in the 3rd group.

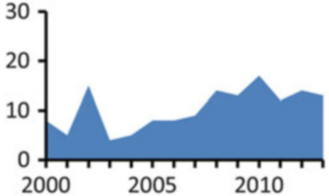
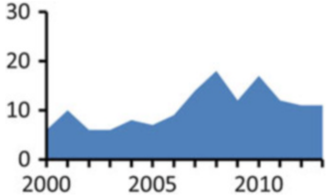
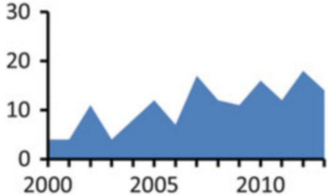
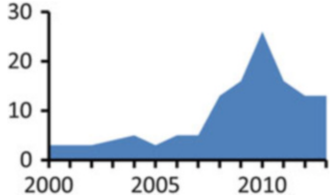
The 1st and 4th groups both have ‘stage’ as the top word and seem to contain several other irrelevant terms. The term ‘stage’ is used widely in the DEA literature, especially in network structure models and contextual analysis methodology. In the context of network structure models, a two-stage process model or two-stage network model refers to a special case of the general network DEA models where a DMU’s operation is divided into two processes and all outputs from the first stage become the only inputs to the second stage. In contextual analysis methodology, two-stage or multi-stage analysis indicates that additional regression analysis is applied to determine the exogenous factors that affect the efficiency. The 1st group includes the term ‘regression’, and so ‘stage’ has to be used in the context of two-stage contextual factor analysis. The 4th group contains terms on new methodologies such ‘slacks-based measure’ (SBM), ‘network DEA’, and ‘dynamic’. Thus, it is reasonable to assume that the term ‘stage’ in this group is related to a two-stage process model. The terms ‘banking’ and ‘financial’ in this group can be interpreted as these new methodologies using banking and financial data to test their new models.

Based on the observations and discussions above, we determine that the four research fronts are “bootstrapping and two-stage analysis”, “undesirable factors”, “cross-efficiency and ranking”, and “network DEA, dynamic DEA, and SBM”. All of them focus on methodologies and techniques. From methodological point of view, it may not be appropriate to group SBM with network and dynamic DEA. They are grouped together here because some of the network and dynamic DEA are based upon SBM.

These four subareas show, nevertheless, dissimilar paper growth trends. The last row in Table 18.3 presents the trend of the number of papers for each subarea. The first three subareas exhibit somewhat jagged growth while the 4th subarea “network DEA, dynamic DEA, and SBM” displays a surge in the period 2008–2012.

A previous study (Liu et al. 2013a) mentions five active subareas—“two-stage contextual factor evaluation framework”, “extending models”, “handling special data”, “examining the internal structure”, and “measuring environmental performance”—using a dataset for the period 1978–2010. In comparison, three of them continue as active subareas: “two-stage contextual factor evaluation framework” (now as “bootstrap and two-stage analysis”), “examining the internal structure”

Table 18.3 Keywords for the major active research subareas

Group ID	1	2	3	4
No. of papers	156	156	152	147
Keywords	Stage (0.67) Bootstrap (0.42) Distributions (0.35) Stochastic (0.31) Banking (0.30) Regression (0.28) Environmental factors (0.13)	Energy (1.15) Electrical power (0.74) Environmental impacts (0.69) Ecological (0.60) Emission (0.44) Undesirable factors (0.40) Policy (0.26) Returns to scale (0.25) Distributions (0.22) CO_2 (0.20) Malmquist (0.13)	Cross efficiency (0.73) Ranking (0.45) Returns to scale (0.14)	Stage (1.05) Banking (0.78) Slacks-based measure (0.46) Network DEA (0.42) Dynamic (0.41) Financial (0.22) Malmquist (0.18) Returns to scale (0.16)
Focus	Bootstrap and two-stage analysis	Undesirable factors	Cross-efficiency and ranking	Network DEA, dynamic DEA, and SBM
Papers by year				

Note: Numbers in the *parentheses* are the total appearance count for keywords in the titles and abstract of all articles within the target group. Only those keywords with an average appearance count greater than 0.12 are listed (now as "network DEA, dynamic DEA, and SBM"), as well as "measuring environmental performance" (now as "undesirable factors"). We elaborate on the contents of these four research fronts in the following subsections.

18.5.1 Bootstrapping and Two-Stage Analysis

DEA does not embed a statistical concept in its original methodology. This particular research stream integrates two statistical methodologies into DEA. The first one, bootstrapping, constructs a base for statistical inference in DEA. The second, two-stage analysis, establishes procedures for contextual factor analysis.

Bootstrapping refers to regenerating original input/output data repeatedly according to a specified statistical distribution. The purpose of the technique is to mimic the sampling distribution of the original data, thus allowing the estimation of data bias and the construction of confidence intervals for efficiency scores. The additional confidence interval information for efficiency scores can be useful for decision makers. The main paths for this research front, shown in Fig. 18.2, begin with SimarW2000 (Simar and Wilson 2000) which extends their earlier work (Simar and Wilson 1998) on bootstrapping in non-parametric frontier models and proposes a general methodology. SimarW2002 (Simar and Wilson 2002) presents another bootstrap procedure, this time for testing hypotheses regarding returns to scale in DEA models. Wilson2003 (Wilson 2003) discusses independence test methods under the premise that an independence condition can simplify the bootstrap procedure.

Finding the contextual factors that affect the efficiency has a strong need for many DEA applications and is the emphasis of many studies. This is usually done through a two-stage procedure that typically begins by calculating DEA efficiency scores and then regressing these scores on contextual factors. As to which regression model is the most proper to use in the second stage is a subject of intense debates. The next study on the main paths, SimarW2007 (Simar and Wilson 2007), suggests that a maximum likelihood estimation of a truncated regression rather than

Fig. 18.2 Key-route main path for "bootstrapping and two-stage analysis" research front (for top 8 key-routes)

tobit is the preferred approach. BankerN2008 (Banker and Natarajan 2008) indicates that ordinary least squares (OLS), maximum likelihood, and tobit regression are all appropriate. Mcdonald2009 (McDonald 2009) advocates using OLS and not using tobit. RamalhoRH2010 (Ramalho et al. 2010) proposes to use fractional regression models. SimarW2011a (Simar and Wilson 2011) compares the approaches in SimarW2007 and BankerN2008 in detail. BadinDS2012 (Badin et al. 2012) suggests a new two-stage type approach for detecting the impact of contextual factors by using a conditional efficiency measure. JohnsonK2012 (Johnson and Kuosmanen 2012) offers a method that directly integrates the regression model into the standard DEA formulation and develop a new one-stage semi-nonparametric estimator.

Two-stage analysis is a useful tool for decision makers who are looking for improving performance while coping with environmental factors. The development in this area in the last few years largely advances our understanding on the assumptions and constraints of the methodology. From a practical point of view, the current state of development, nevertheless, still leaves some confusion to practitioners whose true need is a clear guidance on what methodology to use.

18.5.2 Undesirable Factors

Applying DEA to measure energy and environmental performance faces a special situation where an increase in output level may not be desirable. This is particularly so for some output factors such as wastes, pollutants, and noise. Most of the studies in this subarea focus on the methods that deal with such undesirable outputs. These methods resort to extending into the area of reference technology and/or efficiency measure under the traditional DEA framework (Zhou et al. 2008).

SeifordZ2002 (Seiford and Zhu 2002) and HailuV2001 (Hailu and Veeman 2001) lead the main paths for this research front as shown in Fig. 18.3. HailuV2001 suggest a method equivalent to treating undesirable outputs as inputs. SeifordZ2002, on the other hand, propose to deal with undesirable outputs by applying a monotone decreasing transformation to them. FareG2004 (Fare and Grosskopf 2004) advocates the concept of weak disposability and suggests applying

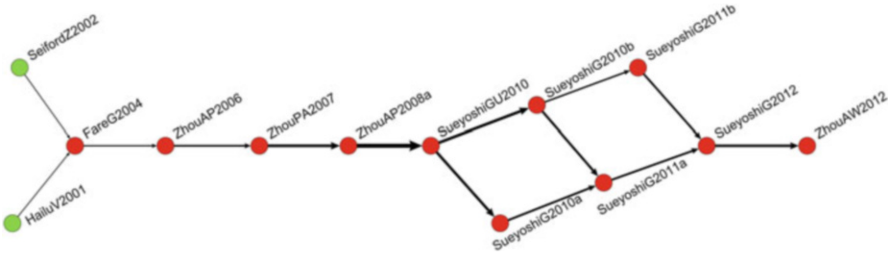


Fig. 18.3 Key-route main path for "undesirable factors" research front (for top 10 key-routes)

a directional distance function efficiency measure. Weak disposability of outputs means that the model mandates a proportional reduction of both desirable and undesirable outputs and that the reduction of only undesirable outputs is impossible. ZhouAP2006 (Zhou et al. 2006) and ZhouPA2007 (Zhou et al. 2007) adopt weak disposability reference technology, but use slacks-based and non-radial measures, respectively, to evaluate the environmental performance of OECD countries.

ZhouAP2008a, as mentioned in Sect. 18.4.2, presents a review article. It is followed by a series of works by Sueyoshi and colleagues (SueyoshiGU2010 (Sueyoshi et al. 2010), SueyoshiG2010a (Sueyoshi and Goto 2010a), SueyoshiG2010b (Sueyoshi and Goto 2010b), SueyoshiG2011a (Sueyoshi and Goto 2011a), SueyoshiG2011b (Sueyoshi and Goto 2011b), and SueyoshiG2012 (Sueyoshi and Goto 2012)). Several of these works are also on the overall main path discussed in Sect. 18.4.2. The most recent study, ZhouAW2012 (Zhou et al. 2012), applies a directional distance function approach to measure energy and CO_2 emission performance in electricity generation.

In summary, the core studies in this subarea evolve around approaches to deal with undesirable output and the attention on the performance of energy and environmental system remains strong as the challenges from energy and pollution have never been more demanding. To further extend the power of modern DEA, applying a network DEA concept to tap into the inner workings of energy and environment systems may be an interesting topic for future research.

18.5.3 Cross-Efficiency and Ranking

The cross-efficiency concept was first proposed by Sexton et al. (1986) in 1986, but did not emerge as a serious alternative DEA approach until Doyle and Green (1994) re-examine it in detail in 1994. The concept increases the discriminating power of DEA by conducting peer-evaluation as opposed to self-evaluation in the traditional DEA model. It is associated closely with the idea of ranking and is widely used in applications where the ranking of DMUs is needed. As shown in Fig. 18.4, the key-route main paths for this subarea begin with a review article on ranking methods.

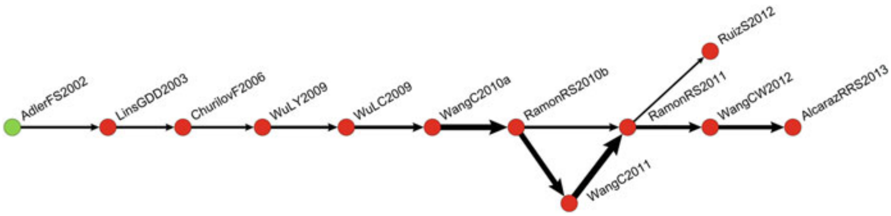


Fig. 18.4 Key-route main path for "cross-efficiency and ranking" research front (for top 10 key-routes)

AdlerFS2002 (Adler et al. 2002) review six ranking methods, including cross-efficiency, under the DEA context.

Four studies following AdlerFS2002 all share an interest in ranking the performance of countries in the Olympics. LinsGDD2003 (Lins et al. 2003) proposes a zero-sum gains model. ChurilovF2006 (Churilov and Flitman 2006) integrates data mining techniques to DEA. WuLY2009 (Wu et al. 2009a) and WuLC2009 (Wu et al. 2009b) apply a cross-efficiency model, while the latter propose a DEA game cross-efficiency model where each DMU is viewed as a competitor via a non-cooperative game.

The remaining parts of the main paths are dominated by articles that propose alternative approaches to improve the cross-efficiency analysis. A core issue in cross-efficiency evaluation is that it may give multiple efficiency scores resulting from alternate optima in solving the linear programming model. Doyle and Green suggest solving this non-uniqueness problem with the use of secondary goals, which include two alternatives called benevolent and aggressive formulations. These secondary goals are related to weight determination. Different from benevolent and aggressive formulations, WangC2010a (Wang and Chin 2010) and WangCW2012 (Wang et al. 2012) propose formulations that determine the input and output weights in a neutral way. RamonRS2010b (Ramon et al. 2010) and RamonRS2011 (Ramon et al. 2011) move further along the idea of WangC2010a. WangC2011 (Wang and Chin 2011) offers a study on the aggregation process in calculating cross-efficiency. RuizS2012 (Ruiz and Sirvent 2012) suggest calculating the cross-efficiency scores using a weighted average rather than an arithmetic mean. AlcarazRRS2013 (Alcaraz et al. 2013) put forward a method that has no need to choose a weighting method and yield a range of possible rankings for each DMU.

As shown in Table 18.3, keywords in this subarea center on only three terms. In fact, it is a truly very focused subarea. Such a large coherent block of research studies indicates that many issues in cross-efficiency remain to be resolved and that there probably has not been a consensus on the method to address the issues in the original cross-efficiency concept. For example, a recent review on cross-efficiency (Cook and Zhu forthcoming) discusses two other alternative approaches including the game cross efficiency methodology of Liang et al. (2008) and the maximum cross efficiency concept of Cook and Zhu (2013).

18.5.4 Network DEA, Dynamic DEA, and SBM

Another chunk of coherent literature consists of several fast evolving DEA topics: SBM, network DEA, and dynamic DEA. There is no surprise that network DEA and dynamic DEA are assigned to the same group as they are conceptually associated with each other. Nevertheless, SBM is grouped with network and dynamic DEA for the reason that some of the network and dynamic DEA are based upon SBM.

The term 'network DEA' was first used in Färe and Grosskopf (2000) in 2000. This work is the most likely candidate for the leading articles on the main paths of

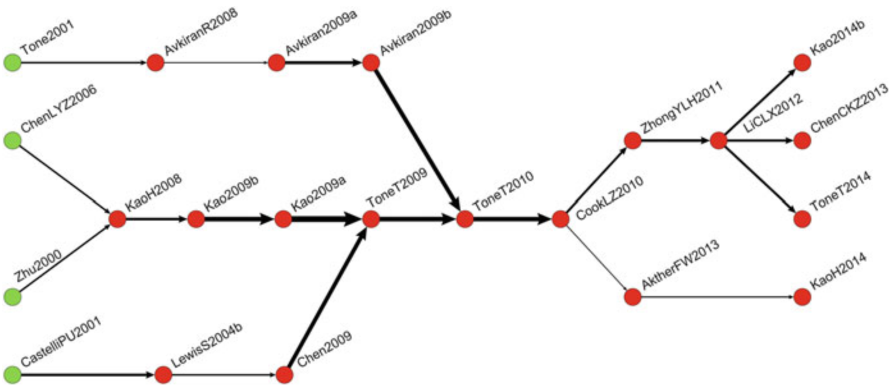


Fig. 18.5 Key-route main paths for "network DEA, dynamic DEA, and SBM" research front (for top 10 key-routes)

this research front. Due to a limitation[6] of this current study, it is not shown on the main paths. In the summary section of the article, Färe and Grosskopf notes: "The basic idea of the network model is to 'connect' processes—providing a single model framework for multi-stage production (with intermediate products, for example) or multi-period production" (Färe and Grosskopf 2000) wherein 'multi-stage production' and 'multi-period production' can be regarded as the origin of network DEA and dynamic DEA concepts, respectively.

The main paths for this group, as shown in Fig. 18.5, begin with three research streams that eventually merge to Tone's work on dynamic DEA, ToneT2010 (Tone and Tsutsui 2010), and then divide into two research activities. Tone2001 (Tone 2001) leads the first research stream, introducing SBM. SBM, as opposed to the radial measure used in the traditional CCR (Charnes et al. 1978) and BCC (Banker et al. 1984) models, measures efficiency based on the input excesses and output shortfalls. AvkiranR2008 (Avkiran and Rowlands 2008) and Avkiran2009a (Avkiran 2009b) extend a three-stage procedure proposed in Fried et al. (2002) by infusing SBM into the procedure. The procedure accounts for environmental effects and statistical noise in the efficiency measure. Avkiran2009b (Avkiran 2009a), in parallel with ToneT2009 (Tone and Tsutsui 2009), proposes a model for network SBM.

The second stream begins with Zhu2000 (Zhu 2000) and ChenLYZ2006 (Chen et al. 2006). Zhu2000 suggests measuring bank performance in two process stages. ChenLYZ2006 proposes a two-stage process model that is an improvement over a model they propose earlier (Chen and Zhu 2004). The core of this stream, however,

[6] The article is published in the Socio-economic Planning Sciences journal, which is not listed in the WOS database. Thus, no citation information is available for the article.

is a series of studies by Kao that propose several relational models (KaoH2008 (Kao and Hwang 2008), Kao2009a (Kao 2009a), and Kao2009b (Kao 2009b)). ToneT2009 (Tone and Tsutsui 2009) extends the network DEA model to the SBM framework.

In the third research stream, CastelliPU2001 (Castelli et al. 2001) presents a model that considers the case of specialized and interdependent subunits in a DMU. LewisS2004b (Lewis and Sexton 2004) proposes a network DEA model with a different reference set definition than that proposed in Färe and Grosskopf (2000). Chen2009 (Chen 2009) incorporates dynamic effects into a network DEA model.

The merging article, ToneT2010 (Tone and Tsutsui 2010), extends the dynamic DEA concept proposed by Färe et al. (1996) within the SBM framework. CookLZ2010 (Cook et al. 2010a) reviews the existing two-stage process models. The remaining studies on the main paths, except ZhongYLH2011 (Zhong et al. 2011), are all the latest studies on network DEA or dynamic DEA. LiCLX2012 (Li et al. 2012) extend the two-stage network structure by allowing inputs to the second stage to come from sources other than the outputs of the first stage. AktherFW2013 (Akther et al. 2013) applies two-stage process model to study bank efficiencies. Kao2014b (Kao 2014a) discusses a general multi-stage DEA model. The model defines efficiency in a different way than that defined by Cook et al. (2010b). ChenCKZ2013 (Chen et al. 2013b) discusses the differences between the multiplier and envelopment network DEA models and points out the functions of each. ToneT2014 (Tone and Tsutsui 2014) offers a model that combines the network DEA and dynamic DEA models under the SBM framework. KaoH2014 (Kao and Hwang 2014) proposes a multi-period two-stage DEA model.

The research activities in this subarea have moved at a very fast pace since 2008. One indication mentioned earlier is that the papers in this subarea have surged during the period 2008–2012. Another indication is the number of review articles. As of 2014, there are already three review papers for this subarea (Castelli et al. 2010; Cook et al. 2010a; Kao 2014b). The most recent review by Kao (2014b) indicates several future research directions for this subarea, including the type of data used, the Malmquist index for network systems, and dynamic analysis of network systems. Incorporating a time factor into the network structure is clearly the core research activity in this research front.

18.5.5 Other Coherent Subareas

In addition to the four research fronts, six smaller coherent research subareas deserve brief notes. Their group sizes are 97, 55, 47, 47, 38, and 32 respectively. For these groups, we list papers with relative high citation counts to highlight their subject of interests. We remind readers that the discussions herein include only papers in Part II data.

The 5th group focuses on banking application. Banking had been at the top of DEA applications over the period 1978–2010 as indicated in Liu et al. (2013b). In this current study, many of the papers examining bank data are clustered into other groups due to the fact that their emphases are on methods and techniques. Cook et al. (2000), Cook and Hababou (2001), Paradi and Schaffnit (2004), and Paradi et al. (2011) apply various techniques to examine the performance of bank branches. Asmild et al. (2004) evaluate Canadian bank performances over time by combining DEA window analysis with the Malmquist index. Two review articles (Fethi and Pasiouras 2010; Paradi and Zhu 2013) are good references on the progress of research in applying DEA to measure bank performances.

The 6th group emphasizes on fuzzy and imprecise DEA. Traditional DEA assumes that input and output data are crisp and precise, but the assumption may not always be true in the real world. Data can be fuzzy (vague, fluctuate), or it can be imprecise (bounded, ordinal, or ratio bounded). Kao and Liu (2000), Guo and Tanaka (2001), Lertworasirikul et al. (2003), and Wang et al. (2005) propose different approaches to handle fuzzy input and output data. Hatami-Marbini et al. (2011) survey and classify fuzzy DEA methods. Cooper et al. (2001) illustrate the use of imprecise DEA. Zhu (2003) reviews the methods to handle imprecise data.

The next group centers on profit and cost efficiency studies. Most of the studies in this group propose methods or models to deal with various profit and cost efficiency measurement situations, including price under the most and least favorable scenarios (Camanho and Dyson 2005), the law of one price (Kuosmanen et al. 2006), incomplete price information (Kuosmanen and Post 2001), and price depending on negotiation (Camanho and Dyson 2008).

Centralized resource allocation is the main subject of the 8th group. There can be situations where a centralized decision maker supervises all DMUs, such that maximizing overall efficiency across units becomes one of his objectives in addition to maximizing individual units. Beasley (2003), Lozano and Villa (2004), Asmild et al. (2009), and Lozano et al. (2004) propose models that cope with such situations.

The main theme of the 9th group is variable selection. DEA results are sensitive to the number of input/output variables. Several papers in this group propose methods and procedures to make the best selection of variables (Cinca and Molinero 2004; Jenkins and Anderson 2003; Pastor et al. 2002; Wagner and Shimshak 2007), or to cope with a large number of variables (Meng et al. 2008).

The 10th group is about models that handle stochastic data. Cooper et al. (2002) propose to apply chance constrained programming to deal with stochastic data. Ruggiero (2004) and Kao and Liu (2009) work on problems with stochastic data using a simulation technique. Dyson and Shale (2010) discuss approaches to handle uncertainty in DEA which includes bootstrapping, Monte Carlo simulation, and change constrained DEA.

18.6 DEA Applications

DEA started out as a theoretical method and found its way into a broad spectrum of applications. In terms of volume, the research papers that apply DEA are now well exceeding those developing theoretical models. In this section, we present several facts related to DEA applications in the light to familiar the readers with the buildup in applications and their association with theoretical models.

A clarification on the meaning of ‘application’ herein is nevertheless necessary before proceeding further. Many DEA papers touch on both methodologies and real world applications. There are basically three types of DEA papers: purely-methodological, application-centered, and theory-together-with-empirical-data (Liu et al. 2013b). The first type, purely-methodological, elaborates on mathematics and models, but does not relate to empirical data, although occasionally some simulated data are used to test the theory. Examples are Banker et al. (1984), who present only mathematics, and Tone (2001), who illustrates the proposed model with a set of artificial data. The second type, application-centered, applies an already developed approach to a real world problem. The focus is mainly on application. Examples are Karkazis and Thanassoulis (1998) and Ma et al. (2002), who apply existing DEA models to study the efficiencies of Greece’s public investment and China’s iron and steel industry, respectively. In-between the two extremes is the third type, theory-together-with-empirical-data. This type proposes a methodological innovation and then validates or tests the proposed method with a set of empirical data. It may put strong emphasis on the application or simply adopt a previous empirical data to test the model. Examples for the former case are Sahoo and Tone (2009) and Kao and Hwang (2008), whereas the latter include Cook and Zhu (2006) and Chen et al. (2006). The contribution to the theory also varies widely in these studies. Practically, it is not easy to differentiate between the second- and the third-type works as there is a wide spectrum on how the authors balance the weight between the models and applications. Herein, as in Liu et al. (2013b), both of the application-centered and theory-together-with-empirical-data are regarded as application-embedded papers, or simply application papers.

In the remaining part of this section, we begin with presenting the evolution of DEA applications. It is followed by a discussion of prevailing DEA applications. In the end, we discuss the association between DEA methods and applications.

18.6.1 Catching Up of DEA Applications

Liu et al. (2013b) analyze their data collected in August, 2010 and find that among all DEA papers, 36.5 % are purely-methodological and 63.5 % are application-embedded, or roughly one-third purely-methodological and two-third application-embedded. They also point out that this one-to-two ratio between types is not how it was during the early stage of the DEA evolution. In the first 20 years of DEA

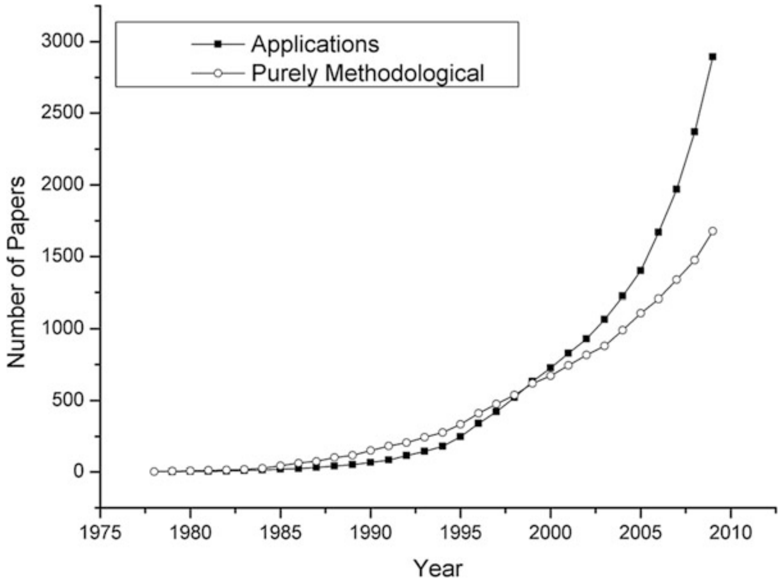


Fig. 18.6 Accumulated number of purely theoretical and application DEA papers (Source: Liu et al. 2013b)

development, purely-methodological articles outnumbered application-embedded papers. It was not until 1999 that the accumulated number of application-embedded papers caught up to the number of purely-methodological papers. Figure 18.6 shows the growth trend of both categories.

18.6.2 Prevailing DEA Application

As regards to prevailing DEA applications, Liu et al. (2013b) suggest that as of 2010 the top five applications are banking, health care, agriculture and farm, transportation, and education. These five applications make up 41.0 % of all application-embedded papers.

It is interesting to know if the emphasis on applications changes with time. In contrast to manually classify the papers as did in Liu et al. (2013b), this study conducts an automatic keyword analysis, and on a fairly recent data—the Part II data. In these papers, only the titles and abstract are examined. We first identify 13 applications[7] and their associated terms. The associated terms are the words or phrases that hint the use of empirical data on certain application. For example,

[7] These 13 applications are banking, health care, agriculture and farm, transportation, education, (electrical) power, manufacturing, energy and environment, communication, finance, insurance, tourism, and fishery.

Table 18.4 Top DEA applications

Rank	Applications	Number of papers[a]	Rank in 2010 data[b]
1	Banking	170	1
2	Energy & environment	130	8
3	Education	125	5
4	Health care	95	2
5	Transportation	77	4
6	Finance	68	10
7	Agriculture & farm	59	3
8	Power	59	6
9	Tourism	57	12
10	Manufacturing	26	7

[a]Number of papers with target application terms in the titles and abstracts of articles in the Part II data
[b]Taken from Table 1 of Liu et al. (2013b)

emission, pollution, energy efficiency, environmental performance, etc. are the terms related to energy and environment application whereas hospital, medical, pharmaceutical, etc. are related to health care application. Whenever a paper contain at least one application specific terms, it is labeled as an article for that application. Papers may be assigned two or more applications if they contains terms for more than one application. Finally, one counts the number of papers for each application.

The result is presented in Table 18.4, which shows scholar's preference of applications in the period 2000–2014. The table is ordered according to the number of papers. The application that attracts the most attention is still banking, which is followed by energy and environment, education, health care, and transportation. The remaining five applications are finance, agriculture and farm, power, tourism, and manufacturing. Notably, the rank for energy and environment application has jumped sharply from number eight to the second.

18.6.3 Association Between Methodologies and Applications

Liu et al. (2013b) examine how methodological works were used in banking, health care, agriculture and farm, transportation, and education. It was done by counting the citations some major methodological papers received by these five application papers. No obvious preferences for each application were observed but there is a general trend of citing network DEA and two-stage analysis modeling articles. It should be noted that this results reflect the status in and before 2010. Fast development of new methodologies and some applications in recent years, may have changed the adoption of models in different applications.

This study re-examine the issue by applying keyword analysis to the recent data (Part II data). We check the usage of application terms in each of the four research

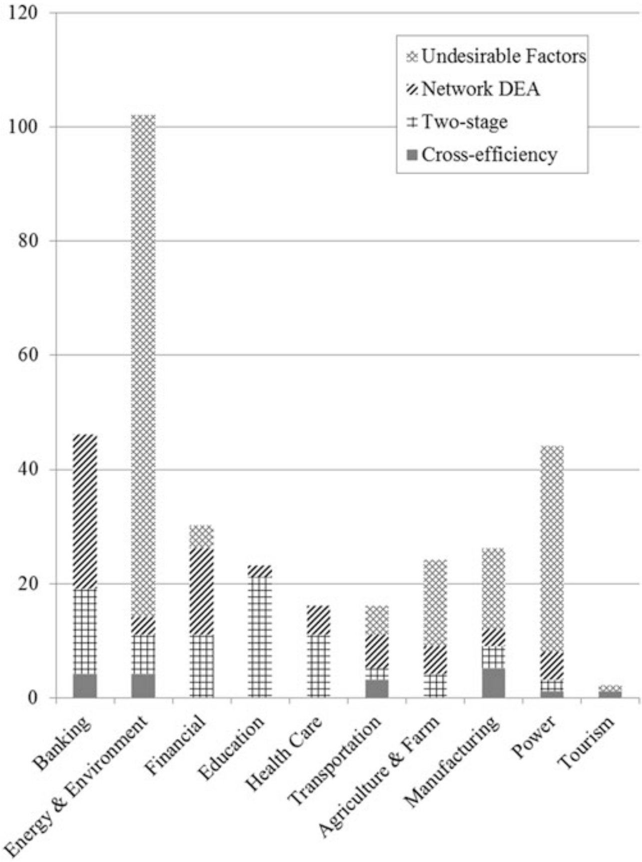


Fig. 18.7 Association between methodologies and applications

fronts, noticing that all these four research fronts focus on methodologies and techniques. In other words, we compare the number of times certain applications are ascribed in each of the four recently popular methodologies. The results are presented in Fig. 18.7. In contrast to the previous results, the preferences of some of the applications on certain methodologies are obvious. Energy and environment as well as power applications largely apply undesirable factor models. Agriculture and farm as well as manufacturing applications are also dominated by undesirable factor model. The results make sense because all these four applications have bad outputs to deal with. Banking and financial applications are associated largely with network DEA and SBM models. Education and health care applications prefer to adopt two-stage analysis. Transportation application does not seem to have clear preference on models while tourism application is not much associated with the four research fronts.

The results of this association analysis imply wide-open research opportunities. Certain preferences of methodologies are rooted in the characteristics of applications. For example, energy and environment applications contain bad outputs thus it is natural to adopt models that are able to handle undesirable factors. Yet many research front methodologies can be applied universally. For example, two-stage analysis is suitable for determining the contextual factors that affect efficiency in all types of application. Network DEA is quite appropriate for examining a process's efficiency in more details. The new methodologies provide opportunities for deeper analyses but many applications have not taken full opportunities from these new methodologies. A caveat is that whether such opportunities really are able to provide meaningful results are up to the judgment of experts in the field.

18.7 Conclusions

The large amount of DEA literature has made it difficult to conduct general surveys using qualitative approaches, thus increasing the need for applying quantitative and systematic approaches in surveying scientific and technological fields. Citation-based analysis, although widely accepted in the research community, has drawn some criticisms, including no discrimination on the level of citation relevancy, self-citation, remote citation, etc. Remote citation is the situation where an article references others in a very broad sense, such as the same application area, the same general method, or even just because of applying the same methodology. Researchers have proposed methods to address some of these shortcomings—for example, Liu et al. (2014) propose methods to handle citations with different levels of relevancy. In addition to issues in citation-based methodology, data in this study are taken only from WOS. Some DEA articles of certain importance may be ignored. Bearing these limitations in mind, we present the research fronts in DEA for the period 2000–2014.

We identifies four research fronts: "bootstrapping and two-stage analysis", "undesirable factors", "cross-efficiency and ranking", and "network DEA, dynamic DEA, and SBM". All of them focus on methodologies and techniques. From the science and technology evolution point of view, two of the research fronts seem to be at a stage that seeks a "dominant design" (Anderson and Tushman 1990; Utterback and Abernathy 1975). The intense debate in the two-stage analysis subarea, over which is the best regression model to adopt, seems to have not yet reached a conclusion. For the network DEA and dynamic DEA subareas, an intellectual exchange that looks for a simple universal model remains to be expected.

The top five applications in the period 2000–2014 are banking, energy and environment, education, health care, and transportation. Growth in energy and environment application is in particular high. Some applications have not incorporated methodologies introduced in recent years. Research opportunities are wide open considering that there are many such gaps to close.

This study contributes to both the DEA and the bibliometric fields. For the former, we present the most recent research fronts that help increase our understanding on those research activities that have continued to make this 35-year old methodology a solid scientific field. It should be noted that the main paths, similar to citation networks, are dynamic. New articles, along with the articles they reference, are changing the main paths day in and day out. What we have presented herein is only a snap shot at the current point in time. As for the contribution to the bibliometric field, the methodology in this study that combines the clustering method and key-route main path analysis turns out to be very effective in exploring research fronts and can be used as a model for future studies targeting research fronts, in DEA, or any other scientific and technology field.

Acknowledgements The authors would like to express their gratitude Professor Joe Zhu for providing guidance and critiques in several aspect of this research project. Our gratitude extends to several reviewers for their constructive comments, which have greatly improved the accuracy and readability of this article. This work is partially supported by Taiwan's Ministry of Science and Technology grant NSC 102-2410-H-011-008-MY3, NSC 102-2410-H-155-021-MY3, and NSC 102-2632-H-155-001-MY3.

Part of this chapter is based upon "John S. Liu, Louis Y.Y. Lu, and Wen-Min Lu, "Research fronts in data envelopment analysis", OMEGA, Vol. 51, (2016), 33–45.", with permission from Elsevier.

References

Adler N, Friedman L, Sinuany-Stern Z (2002) Review of ranking methods in the data envelopment analysis context. Eur J Oper Res 140:249–265

Akther S, Fukuyama H, Weber WL (2013) Estimating two-stage network slacks-based inefficiency: an application to Bangladesh banking. OMEGA Int J Manag Sci 41:88–96

Alcaraz J, Ramon N, Ruiz J, Sirvent I (2013) Ranking ranges in cross-efficiency evaluations. Eur J Oper Res 226:516–521

Anderson P, Tushman ML (1990) Technological discontinuities and dominant designs: a cyclical model of technological change. Adm Sci Q 35:604–633

Angulo-Meza L, Lins MPE (2002) Review of methods for increasing discrimination in data envelopment analysis. Ann Oper Res 116:225–242

Asmild M, Paradi J, Aggarwall V, Schaffnit C (2004) Combining DEA window analysis with the Malmquist index approach in a study of the Canadian banking industry. J Prod Anal 21:67–89

Asmild M, Paradi JC, Pastor JT (2009) Centralized resource allocation bcc models. OMEGA Int J Manag Sci 37:40–49

Avkiran N (2009a) Opening the black box of efficiency analysis: an illustration with UAE banks. OMEGA Int J Manag Sci 37:930–941

Avkiran N (2009b) Removing the impact of environment with units-invariant efficient frontier analysis: an illustrative case study with intertemporal panel data. OMEGA Int J Manag Sci 37:535–544

Avkiran NK, Rowlands T (2008) How to better identify the true managerial performance: state of the art using DEA. OMEGA Int J Manag Sci 36:317–324

Badin L, Daraio C, Simar L (2012) How to measure the impact of environmental factors in a nonparametric production model. Eur J Oper Res 223:818–833

Banker R, Natarajan R (2008) Evaluating contextual variables affecting productivity using data envelopment analysis. Oper Res 56:48–58
Banker RD, Charnes A, Cooper WW (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis. Manag Sci 30:1078–1092
Batagelj V, Mrvar A (1998) Pajek-program for large network analysis. Connections 21:47–57
Beasley JE (2003) Allocating fixed costs and resources via data envelopment analysis. Eur J Oper Res 147:198–216
Berger AN, Humphrey DB (1997) Efficiency of financial institutions: international survey and directions for future research. Eur J Oper Res 98:175–212
Camanho AS, Dyson RG (2005) Cost efficiency measurement with price uncertainty: a DEA application to bank branch assessments. Eur J Oper Res 161:432–446
Camanho A, Dyson R (2008) A generalisation of the Farrell cost efficiency measure applicable to non-fully competitive settings. OMEGA Int J Manag Sci 36:147–162
Castelli L, Pesenti R, Ukovich W (2001) DEA-like models for efficiency evaluations of specialized and interdependent units. Eur J Oper Res 132:274–286
Castelli L, Pesenti R, Ukovich W (2010) A classification of DEA models when the internal structure of the decision making units is considered. Ann Oper Res 173:207–235
Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision making units. Eur J Oper Res 2:429–444
Chen C-M (2009) A network-DEA model with new efficiency measures to incorporate the dynamic effect in production networks. Eur J Oper Res 194:687–699
Chen Y, Zhu J (2004) Measuring information technology's indirect impact on firm performance. Inf Technol Manag 5:9–22
Chen Y, Liang L, Yang F, Zhu J (2006) Evaluation of information technology investment: a data envelopment analysis approach. Comput Oper Res 33:1368–1379
Chen Y-B, Liu JS, Lin P (2013a) Recent trend in graphene for optoelectronics. J Nanopart Res 15:1–14
Chen Y, Cook WD, Kao C, Zhu J (2013b) Network DEA pitfalls: divisional efficiency and frontier projection under general network structures. Eur J Oper Res 226:507–515
Chen Y, Du J, Huo J (2013c) Super-efficiency based on a modified directional distance function. OMEGA Int J Manag Sci 41:621–625
Chuang TC, Liu JS, Lu LY, Lee Y (2014) The main paths of medical tourism: from transplantation to beautification. Tour Manage 45:49–58
Churilov L, Flitman A (2006) Towards fair ranking of Olympics achievements: the case of Sydney 2000. Comput Oper Res 33:2057–2082
Cinca CS, Molinero CM (2004) Selecting DEA specifications and ranking units via PCA. J Oper Res Soc 55:521–528
Cook W, Hababou M (2001) Sales performance measurement in bank branches. OMEGA Int J Manag Sci 29:299–307
Cook WD, Seiford LM (2009) Data envelopment analysis (DEA)–thirty years on. Eur J Oper Res 192:1–17
Cook WD, Zhu J (2006) Rank order data in DEA: a general framework. Eur J Oper Res 174:1021–1038
Cook W, Zhu J (2013) DEA Cobb-Douglas frontier and cross-efficiency. J Oper Res Soc 65:265–268
Cook W, Hababou M, Tuenter H (2000) Multicomponent efficiency measurement and shared inputs in data envelopment analysis: an application to sales and service performance in bank branches. J Prod Anal 14:209–224
Cook W, Liang L, Zhu J (2010a) Measuring performance of two-stage network structures by DEA: a review and future perspective. OMEGA Int J Manag Sci 38:423–430
Cook WD, Zhu J, Bi G, Yang F (2010b) Network DEA: additive efficiency decomposition. Eur J Oper Res 207:1122–1129

Cook WD, Tone K, Zhu J (2014) Data envelopment analysis: prior to choosing a model. OMEGA Int J Manag Sci 44:1–4

Cook WD, Zhu J (2015) DEA cross efficiency. In: Zhu Joe (ed) Data envelopment analysis: a handbook of models and methods. Springer US

Cooper WW, Park KS, Yu G (2001) An illustrative application of idea (imprecise data envelopment analysis) to a Korean mobile telecommunication company. Oper Res 49:807–820

Cooper WW, Deng H, Huang Z, Li SX (2002) Chance constrained programming approaches to technical efficiencies and inefficiencies in stochastic data envelopment analysis. J Oper Res Soc 53:1347–1356

Cooper W, Seiford LM, Tone K, Zhu J (2007) Some models and measures for evaluating performances with DEA: past accomplishments and future prospects. J Prod Anal 28:151–163

Csardi G, Nepusz T (2006) The igraph software package for complex network research. InterJournal, Complex Systems 1695

Dai X, Kuosmanen T (2014) Best-practice benchmarking using clustering methods: application to energy regulation. OMEGA Int J Manag Sci 42:179–188

Doyle J, Green R (1994) Efficiency and cross-efficiency in DEA: derivations, meanings and uses. J Oper Res Soc 567–578

Dyson R, Shale E (2010) Data envelopment analysis, operational research and uncertainty. J Oper Res Soc 61:25–34

Egghe L (2006) Theory and practise of the g-index. Scientometrics 69:131–152

Emrouznejad A, Parker BR, Tavares G (2008) Evaluation of research in efficiency and productivity: a survey and analysis of the first 30 years of scholarly literature in DEA. Socioecon Plann Sci 42:151–157

Fang H, Lee H, Hwang S, Chung C (2013) A slacks-based measure of super-efficiency in data envelopment analysis: an alternative approach. OMEGA Int J Manag Sci 41:731–734

Färe R, Grosskopf S (2000) Network DEA. Socioecon Plann Sci 34:35–49

Fare R, Grosskopf S (2004) Modeling undesirable factors in efficiency evaluation: comment. Eur J Oper Res 157:242–245

Färe R, Grosskopf S, Brännlund R (1996) Intertemporal production frontiers: with dynamic DEA. Kluwer, Boston

Fethi MD, Pasiouras F (2010) Assessing bank efficiency and performance with operational research and artificial intelligence techniques: a survey. Eur J Oper Res 204:189–198

Fried HO, Lovell CK, Schmidt SS, Yaisawarng S (2002) Accounting for environmental effects and statistical noise in data envelopment analysis. J Prod Anal 17:157–174

Gattoufi S, Oral M, Kumar A, Reisman A (2004a) Epistemology of data envelopment analysis and comparison with other fields of OR/MS for relevance to applications. Socioecon Plann Sci 38:123–140

Gattoufi S, Oral M, Reisman A (2004b) Data envelopment analysis literature: a bibliography update (1951–2001). J Socioecon Plann Sci 38:159–229

Girvan M, Newman ME (2002) Community structure in social and biological networks. Proc Natl Acad Sci U S A 99:7821–7826

Guo P, Tanaka H (2001) Fuzzy DEA: a perceptual evaluation method. Fuzzy Set Syst 119:149–160

Hailu A, Veeman T (2001) Non-parametric productivity analysis with undesirable outputs: an application to the Canadian pulp and paper industry. Am J Agric Econ 83:605–616

Hatami-Marbini A, Emrouznejad A, Tavana M (2011) A taxonomy and review of the fuzzy data envelopment analysis literature: two decades in the making. Eur J Oper Res 214:457–472

Hirsch JE (2005) An index to quantify an individual's scientific research output. Proc Natl Acad Sci U S A 102:16569–16572

Hosseinzadeh Lotfi F, Jahanshahloo GR, Khodabakhshi M, Rostamy-Malkhlifeh M, Moghaddas Z, Vaez-Ghasemi M (2013) A review of ranking models in data envelopment analysis. J Appl Math 2013

Hummon NP, Dereian P (1989) Connectivity in a citation network: the development of DNA theory. Soc Networks 11:39–63

Hung S-C, Liu JS, Lu LY, Tseng Y-C (2014) Technological change in lithium iron phosphate battery: the key-route main path analysis. Scientometrics 1–24

Jenkins L, Anderson M (2003) A multivariate statistical approach to reducing the number of variables in data envelopment analysis. Eur J Oper Res 147:51–61

Johnson A, Kuosmanen T (2012) One-stage and two-stage DEA estimation of the effects of contextual variables. Eur J Oper Res 220:559–570

Kao C (2009a) Efficiency decomposition in network data envelopment analysis: a relational model. Eur J Oper Res 192:949–962

Kao C (2009b) Efficiency measurement for parallel production systems. Eur J Oper Res 196:1107–1112

Kao C (2014a) Efficiency decomposition for general multi-stage systems in data envelopment analysis. Eur J Oper Res 232:117–124

Kao C (2014b) Network data envelopment analysis: a review. Eur J Oper Res 239:1–16

Kao C, Hwang S (2008) Efficiency decomposition in two-stage data envelopment analysis: an application to non-life insurance companies in Taiwan. Eur J Oper Res 185:418–429

Kao C, Hwang S-N (2014) Multi-period efficiency and Malmquist productivity index in two-stage production systems. Eur J Oper Res 232:512–521

Kao C, Liu S-T (2000) Fuzzy efficiency measures in data envelopment analysis. Fuzzy Set Syst 113:427–437

Kao C, Liu S-T (2009) Stochastic data envelopment analysis in measuring the efficiency of Taiwan commercial banks. Eur J Oper Res 196:312–322

Karkazis J, Thanassoulis E (1998) Assessing the effectiveness of regional development policies in northern Greece using data envelopment analysis. Socioecon Plann Sci 32:123–137

Kuosmanen T, Post T (2001) Measuring economic efficiency with incomplete price information: with an application to European commercial banks. Eur J Oper Res 134:43–58

Kuosmanen T, Cherchye L, Sipiläinen T (2006) The law of one price in data envelopment analysis: restricting weight flexibility across firms. Eur J Oper Res 170:735–757

Lertworasirikul S, Fang S-C, Joines JA, Nuttle HL (2003) Fuzzy data envelopment analysis (DEA): a possibility approach. Fuzzy Set Syst 139:379–394

Lewis HF, Sexton TR (2004) Network DEA: efficiency analysis of organizations with complex internal structure. Comput Oper Res 31:1365–1410

Li Y, Chen Y, Liang L, Xie J (2012) DEA models for extended two-stage network structures. OMEGA Int J Manag Sci 40:611–618

Liang L, Wu J, Cook WD, Zhu J (2008) The DEA game cross-efficiency model and its Nash equilibrium. Oper Res 56:1278–1288

Lins MPE, Gomes EG, Soares de Mello JCC, Soares de Mello AJR (2003) Olympic ranking based on a zero sum gains DEA model. Eur J Oper Res 148:312–322

Liu JS, Lu LY (2012) An integrated approach for main path analysis: development of the Hirsch index as an example. J Am Soc Inf Sci Technol 63:528–542

Liu JS, Lu LYY, Lu WM, Lin BJY (2013a) Data envelopment analysis 1978–2010: a citation-based literature survey. OMEGA Int J Manag Sci 41:3–15

Liu JS, Lu LYY, Lu WM, Lin BJY (2013b) A survey of DEA applications. OMEGA Int J Manag Sci 41:893–902

Liu JS, Chen HH, Ho MHC, Li YC (2014) Citations with different levels of relevancy: tracing the main paths of legal opinions. J Assoc Inf Sci Technol 65:2479–2488

Lozano S, Villa G (2004) Centralized resource allocation using data envelopment analysis. J Prod Anal 22:143–161

Lozano S, Villa G, Adenso-Dıaz B (2004) Centralised target setting for regional recycling operations using DEA. OMEGA Int J Manag Sci 32:101–110

Lu LY, Liu JS (2013) An innovative approach to identify the knowledge diffusion path: the case of resource-based theory. Scientometrics 94:225–246

Lu LY, Liu JS (2014) The knowledge diffusion paths of corporate social responsibility–from 1970 to 2011. Corp Soc Responsib Environ Manag 21:113–128
Ma J, Evans DG, Fuller RJ, Stewart DF (2002) Technical efficiency and productivity change of china's iron and steel industry. Int J Prod Econ 76:293–312
McDonald J (2009) Using least squares and tobit in second stage DEA efficiency analyses. Eur J Oper Res 197:792–798
Meng W, Zhang D, Qi L, Liu W (2008) Two-level DEA approaches in research evaluation. OMEGA Int J Manag Sci 36:950–957
Mukherjee A, Nath P, Pal M (2003) Resource, service quality and performance triad: a framework for measuring efficiency of banking services. J Oper Res Soc 54:723–735
Newman ME (2006) Modularity and community structure in networks. Proc Natl Acad Sci U S A 103:8577–8582
Paradi JC, Schaffnit C (2004) Commercial branch performance evaluation and results communication in a Canadian bank—a DEA application. Eur J Oper Res 156:719–735
Paradi J, Zhu H (2013) A survey on bank branch efficiency and performance research with data envelopment analysis. OMEGA Int J Manag Sci 41:61–79
Paradi J, Rouatt S, Zhu H (2011) Two-stage evaluation of bank branch efficiency using data envelopment analysis. OMEGA Int J Manag Sci 39:99–109
Pastor JT, Ruiz JL, Sirvent I (2002) A statistical test for nested radial DEA models. Oper Res 50:728–735
Ramalho E, Ramalho J, Henriques P (2010) Fractional regression models for second stage DEA efficiency analyses. J Prod Anal 34:239–255
Ramanathan R (2005) Estimating energy consumption of transport modes in India using DEA and application to energy and environmental policy. J Oper Res Soc 56:732–737
Ramon N, Ruiz J, Sirvent I (2010) On the choice of weights profiles in cross-efficiency evaluations. Eur J Oper Res 207:1564–1572
Ramon N, Ruiz J, Sirvent I (2011) Reducing differences between profiles of weights: a "peer-restricted" cross-efficiency evaluation. OMEGA Int J Manag Sci 39:634–641
Ruggiero J (2004) Data envelopment analysis with stochastic data. J Oper Res Soc 55:1008–1012
Ruiz J, Sirvent I (2012) On the DEA total weight flexibility and the aggregation in cross-efficiency evaluations. Eur J Oper Res 223:732–738
Sahoo BK, Tone K (2009) Decomposing capacity utilization in data envelopment analysis: an application to banks in India. Eur J Oper Res 195:575–594
Seiford LM (1996) Data envelopment analysis: the evolution of the state of the art (1978–1995). J Prod Anal 7:99–137
Seiford LM, Thrall RM (1990) Recent developments in DEA: the mathematical programming approach to frontier analysis. J Econ 46:7–38
Seiford LM, Zhu J (1998a) Sensitivity analysis of DEA models for simultaneous changes in all the data. J Oper Res Soc 49:1060–1071
Seiford LM, Zhu J (1998b) Stability regions for maintaining efficiency in data envelopment analysis. Eur J Oper Res 108:127–139
Seiford LM, Zhu J (1999a) Infeasibility of super-efficiency data envelopment analysis models. INFOR 37:174–187
Seiford LM, Zhu J (1999b) Profitability and marketability of the top 55 us commercial banks. Manag Sci 45:1270–1288
Seiford L, Zhu J (2002) Modeling undesirable factors in efficiency evaluation. Eur J Oper Res 142:16–20
Sexton TR, Silkman RH, Hogan AJ (1986) Data envelopment analysis: critique and extensions. In: Silkman Richard H (ed) Measuring efficiency: an assessment of data envelopment analysis. Jossey-Bass, San Francisco, pp 73–105
Sherman HD, Zhu J (2013) Analyzing performance in service organizations. Sloan Manage Rev 54:36–42

Simar L, Wilson PW (1998) Sensitivity analysis of efficiency scores: how to bootstrap in nonparametric frontier models. Manag Sci 44:49–61
Simar L, Wilson P (2000) A general methodology for bootstrapping in non-parametric frontier models. J Appl Stat 27:779–802
Simar L, Wilson P (2002) Non-parametric tests of returns to scale. Eur J Oper Res 139:115–132
Simar L, Wilson P (2007) Estimation and inference in two-stage, semi-parametric models of production processes. J Econ 136:31–64
Simar L, Wilson P (2011) Two-stage DEA: caveat emptor. J Prod Anal 36:205–218
Song M, An Q, Zhang W, Wang Z, Wu J (2012) Environmental efficiency evaluation based on data envelopment analysis: a review. Renew Sustain Energy Rev 16:4465–4469
Sueyoshi T, Goto M (2010a) Measurement of a linkage among environmental, operational, and financial performance in Japanese manufacturing firms: a use of data envelopment analysis with strong complementary slackness condition. Eur J Oper Res 207:1742–1753
Sueyoshi T, Goto M (2010b) Should the us clean air act include CO2 emission control?: examination by data envelopment analysis. Energy Policy 38:5902–5911
Sueyoshi T, Goto M (2011a) Measurement of returns to scale and damages to scale for DEA-based operational and environmental assessment: how to manage desirable (good) and undesirable (bad) outputs? Eur J Oper Res 211:76–89
Sueyoshi T, Goto M (2011b) Methodological comparison between two unified (operational and environmental) efficiency measurements for environmental assessment. Eur J Oper Res 210:684–693
Sueyoshi T, Goto M (2012) Data envelopment analysis for environmental assessment: comparison between public and private ownership in petroleum industry. Eur J Oper Res 216:668–678
Sueyoshi T, Goto M (2013) Returns to scale vs. damages to scale in data envelopment analysis: an impact of U.S. clean air act on coal-fired power plants. OMEGA Int J Manag Sci 41:164–175
Sueyoshi T, Goto M, Ueno T (2010) Performance analysis of us coal-fired power plants by measuring three DEA efficiencies. Energy Policy 38:1675–1688
Tone K (2001) A slacks-based measure of efficiency in data envelopment analysis. Eur J Oper Res 130:498–509
Tone K (2002) A slacks-based measure of super-efficiency in data envelopment analysis. Eur J Oper Res 143:32–41
Tone K, Tsutsui M (2009) Network DEA: a slacks-based measure approach. Eur J Oper Res 197:243–252
Tone K, Tsutsui M (2010) Dynamic DEA: a slacks-based measure approach. OMEGA Int J Manag Sci 38:145–156
Tone K, Tsutsui M (2014) Dynamic DEA with network structure: a slacks-based measure approach. OMEGA Int J Manag Sci 42:124–131
Utterback JM, Abernathy WJ (1975) A dynamic model of process and product innovation. OMEGA Int J Manag Sci 3:639–656
Wagner JM, Shimshak DG (2007) Stepwise selection of variables in data envelopment analysis: procedures and managerial perspectives. Eur J Oper Res 180:57–67
Wang Y, Chin K (2010) A neutral DEA model for cross-efficiency evaluation and its extension. Expert Syst Appl 37:3666–3675
Wang Y, Chin K (2011) The use of OWA operator weights for cross-efficiency aggregation. OMEGA Int J Manag Sci 39:493–503
Wang Y-M, Greatbanks R, Yang J-B (2005) Interval efficiency assessment using data envelopment analysis. Fuzzy Set Syst 153:347–370
Wang Y, Chin K, Wang S (2012) DEA models for minimizing weight disparity in cross-efficiency evaluation. J Oper Res Soc 63:1079–1088
Wilson P (2003) Testing independence in models of productive efficiency. J Prod Anal 20:361–390
Wu J, Liang L, Chen Y (2009a) DEA game cross-efficiency approach to Olympic rankings. OMEGA Int J Manag Sci 37:909–918

Wu J, Liang L, Yang F (2009b) Achievement and benchmarking of countries at the summer Olympics using cross efficiency evaluation method. Eur J Oper Res 197:722–730

Zhong W, Yuan W, Li S, Huang Z (2011) The performance evaluation of regional R&D investments in china: an application of DEA based on the first official china economic census data. OMEGA Int J Manag Sci 39:447–455

Zhou P, Ang B, Poh K (2006) Slacks-based efficiency measures for modeling environmental performance. Ecol Econ 60:111–118

Zhou P, Poh K, Ang B (2007) A non-radial DEA approach to measuring environmental performance. Eur J Oper Res 178:1–9

Zhou P, Ang B, Poh K (2008) A survey of data envelopment analysis in energy and environmental studies. Eur J Oper Res 189:1–18

Zhou P, Ang B, Wang H (2012) Energy and CO2 emission performance in electricity generation: a non-radial directional distance function approach. Eur J Oper Res 221:625–635

Zhu J (2000) Multi-factor performance measure model with an application to fortune 500 companies. Eur J Oper Res 123:105–124

Zhu J (2001) Super-efficiency and DEA sensitivity analysis. Eur J Oper Res 129:443–455

Zhu J (2003) Imprecise data envelopment analysis (idea): a review and improvement with an application. Eur J Oper Res 144:513–529

Index

J. Zhu (ed.), *Data Envelopment Analysis*, International Series in Operations Research & Management Science 238, DOI 10.1007/978-1-4899-7684-0