Emerald Books

Applications of Management Science
Volume 13

Financial Modeling Applications and Data Envelopment Applications

Kenneth D. Lawrence
Gary Kleinman
Editors

Emerald

FINANCIAL MODELING APPLICATIONS AND DATA ENVELOPMENT APPLICATIONS

APPLICATIONS OF MANAGEMENT SCIENCE

Series Editor: Kenneth D. Lawrence

Recent Volumes:

Volume 10: Multi-Criteria Applications: Edited by Kenneth D. Lawrence, Gary R. Reeves, and Ronald K. Klimberg

Volume 11: Mathematical Programming: Edited by Kenneth D. Lawrence

Volume 12: Applications of Management Science in Productivity, Finance and Operations: Edited by Kenneth D. Lawrence and Ronald K. Klimberg

APPLICATIONS OF MANAGEMENT SCIENCE
VOLUME 13

FINANCIAL MODELING APPLICATIONS AND DATA ENVELOPMENT APPLICATIONS

EDITED BY

KENNETH D. LAWRENCE
School of Management,
New Jersey Institute of Technology, Newark, NJ, USA

GARY KLEINMAN
Graduate School of Business, Touro College,
New York, NY, USA

United Kingdom – North America – Japan
India – Malaysia – China

Emerald Group Publishing Limited
Howard House, Wagon Lane, Bingley BD16 1WA, UK

First edition 2009

Reprints and permission service
Contact: booksandseries@emeraldinsight.com

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-1-84855-878-6
ISSN: 0276-8976 (Series)

Awarded in recognition of Emerald's production department's adherence to quality systems and processes when preparing scholarly journals for print

INVESTOR IN PEOPLE

CONTENTS

SECTION B: DATA ENVELOPMENT ANALYSIS (DEA) APPLICATIONS I

SECTION C: DATA ENVELOPMENT ANALYSIS (DEA) APPLICATIONS II

LIST OF CONTRIBUTORS

Emel Aktaş	Istanbul Technical University, Istanbul, Turkey
Seungkee Baek	University of Michigan, Ann Arbor, MI, USA
Walter A. Garrett, Jr.	Saint Louis University, St. Louis, MO, USA
Robert R. Grauer	Simon Fraser University, Burnaby, BC, Canada
Karen M. Hogan	Saint Joseph's University, Philadelphia, PA, USA
Cesse Ip	Saint Louis University, St. Louis, MO, USA
Özgür Kabak	Istanbul Technical University, Istanbul, Turkey
Andreas Kleine	University of Hohenheim, Stuttgart, Germany
Gary Kleinman	Touro College, New York, NY, USA
Ronald K. Klimberg	Saint Joseph's University, Philadelphia, PA, USA
N. K. Kwak	Saint Louis University, St. Louis, MO, USA
Tanya Lal	Saint Joseph's University, Philadelphia, PA, USA
Kenneth D. Lawrence	New Jersey Institute of Technology, Newark, NJ, USA
Sheila M. Lawrence	Rutgers University, Newark, NJ, USA
Chang Won Lee	Hanyang University, Seoul, South Korea
Harvey Lermack	Philadelphia University

Reuven R. Levary	Saint Louis University, St. Louis, MO, USA
Amy F. Lipton	Saint Joseph's University, Philadelphia, PA, USA
D. K. Malhotra	Philadelphia University, Philadelphia, PA, USA
Rashmi Malhotra	Saint Joseph's University, Philadelphia, PA, USA
Holly Morehouse	Vermont Department of Education, Montpelier, VT, USA
Daniel Mrazik	Saint Joseph's University, Philadelphia, PA, USA
Gerard T. Olson	Villanova University, Villanova, PA, USA
Şule Önsel	Dogus University, Istanbul, Turkey
Dinesh R. Pai	Rutgers University, Newark, NJ, USA
Samuel J. Ratick	Clark University, Worcester, MA, USA
Bartosz Sawik	AGH University of Science and Technology, Kraków, Poland
Regina Schlindwein	University of Hohenheim, Stuttgart, Germany
Warren T. Sutton	University of Michigan, Ann Arbor, MI, USA
Füsun Ülengin	Dogus University, Istanbul, Turkey
Ira Yermish	Saint Joseph's University, Philadelphia, PA, USA

EDITORIAL BOARD

SECTION A
FINANCIAL APPLICATIONS PAPERS

LEXICOGRAPHIC AND WEIGHTING APPROACH TO MULTI-CRITERIA PORTFOLIO OPTIMIZATION BY MIXED INTEGER PROGRAMMING

Bartosz Sawik

ABSTRACT

*This chapter presents the portfolio optimization problem formulated as a multi-criteria mixed integer program. Weighting and lexicographic approach are proposed. The portfolio selection problem considered is based on a single-period model of investment. An extension of the Markowitz portfolio optimization model is considered, in which the variance has been replaced with the Value-at-Risk (*VaR*). The* VaR *is a quantile of the return distribution function. In the classical Markowitz approach, future returns are random variables controlled by such parameters as the portfolio efficiency, which is measured by the expectation, whereas risk is calculated by the standard deviation. As a result, the classical problem is formulated as a quadratic program with continuous variables and some side constraints. The objective of the problem considered in this chapter is to allocate wealth on different securities to maximize the weighted difference of the portfolio expected*

Financial Modeling Applications and Data Envelopment Applications
Applications of Management Science, Volume 13, 3–18

ISSN: 0276-8976/doi:10.1108/S0276-8976(2009)0000013003

return and the threshold of the probability that the return is less than a required level. The auxiliary objectives are minimization of risk probability of portfolio loss and minimization of the number of security types in portfolio. The four types of decision variables are introduced in the model: a continuous wealth allocation variable that represents the percentage of wealth allocated to each asset, a continuous variable that prevents the probability that return of investment is not less than required level, a binary selection variable that prevents the choice of portfolios whose VaR *is below the minimized threshold, and a binary selection variable that represents choice of stocks in which capital should be invested. The results of some computational experiments with the mixed integer programming approach modeled on a real data from the Warsaw Stock Exchange are reported.*

INTRODUCTION

The overall process of selecting a portfolio is divided into two stages (Markowitz, 1952). The first stage starts with observation, experience, and ends with beliefs about the future performances of available securities. The second stage starts with relevant beliefs about future performances and ends with the choice of portfolio. One type of rule concerning choice of portfolio is that the investor should maximize the capitalized value of future returns. The decision maker places all his funds in the security with the greatest discounted value. Investor does diversify his founds among all those securities that give maximum expected return. If two or more securities have the same value, then any of these or any combination of these is as good as any other. However, the portfolio with maximum expected return is not necessarily the one with minimum risk. The law of large numbers will insure that the actual yield of the portfolio will be almost the same as the expected yield. There is a rate at which the investor can gain expected return by taking on risk measure or reduce risk by giving up expected return.

The optimal security selection is a classical portfolio problem since the seminal works of Markowitz (1952,1959). It consists of picking the best amount of securities, with the aim of maximizing future returns. It is a typical multivariate problem: the only way to improve future returns is to increase the risk level that the decision maker is willing to accept.

The portfolio selection problem is usually considered as a bi-criteria optimization problem where a reasonable trade-off between expected rate of return and risk is seeking. The aim is to maximize future returns by picking the best amount of stocks. The only way to improve future returns is to increase the risk level that the decision maker is disposed to accept (Ogryczak, 2000). In the classical Markowitz model, future returns are random variables that can be controlled by the two parameters: a portfolio's efficiency calculated by the expectation and a risk that is measured with variance. The classical problem is formulated as a quadratic program with continuous variables and some side constraints.

Although the original Markowitz model forms a quadratic programming problem, following Sharpe (1971), many attempts have been made to linearize the portfolio optimization procedure (Speranza, 1993). The linear program solvability is very important for applications to real-life financial and other decisions where the constructed portfolios have to meet numerous side constraints. Examples of them are minimum transaction lots, transaction costs, or mutual funds characteristics. The introduction of these features leads to mixed integer program problems.

This chapter presents a multi-criteria extension of the Markowitz portfolio optimization model, in which the variance has been replaced with the Value-at-Risk (*VaR*). The *VaR* is a quantile of the return distribution function (Esch et al., 2005). The advantage of using this measure in portfolio optimization is that this value of risk is independent of any distribution hypothesis (Benati & Rizzi, 2007; Sarykalin, Serraino, & Uryasev, 2008).

This portfolio optimization problem is formulated as a multi-criteria mixed integer program. The portfolio selection problem considered is based on a single-period model of investment. The problem objective is to allocate wealth on different assets to maximize the portfolio expected return, the threshold of the probability that the return is not less than a required level, the amount of wealth to be invested, and minimization of number of stocks in optimal portfolio.

The four types of decision variables are introduced in the model: a continuous wealth allocation variable that represents the percentage of wealth allocated to each asset, a continuous variable that prevents the probability that return of investment is not less than required level, a binary selection variable that prevents the choice of portfolios whose *VaR* is below the minimized threshold and binary selection variable that represents choice of stocks in which capital should be invested. The results of some computational experiments with the mixed integer programming approach modeled after a real data from the Warsaw Stock Exchange are reported.

In computational experiments, the datasets with time series of the daily quotation of returns of stocks from the Warsaw Stock Exchange were used. The seven years time horizon from February 1, 2000, to February 1, 2007, in total 1,758 days was considered. Also the two years time horizon from

February 4, 2005, to February 1, 2007, in total 500 days and 100 days was considered, with the selection of 139 stocks for portfolio, quoted each day in the historical horizon. Probability of realization for expected stocks returns is the same for each day and summed up for whole period to one.

VALUE-AT-RISK

The formal definition of *VaR* is the α-quantile of the return distribution function, $\alpha \in (0, 1)$, where α is usually chosen to be 0.01, 0.05, or 0.10. The *VaR* of the assets in portfolio for the selected time period t and the probability level q is defined as an amount termed *VaR*, so that the variation of selected asset price return observed during the interval $[0; t]$ will only be less than *VaR* with probability of $(1-q)$. The advantage of using *VaR* measure in portfolio optimization is that this value of risk is independent of any distribution hypothesis (Benati & Rizzi, 2007). It concerns only downside risk, namely the risk of loss. This index measures the loss in question in a certain way. Finally *VaR* is valid for all types of assets and therefore either involve the various valuation models or be independent of these models (Esch et al., 2005).

LEXICOGRAPHIC AND WEIGHTING APPROACH

Mathematical programming approach deals with optimization problems of maximizing or minimizing a function of many variables subject to inequality and equality constraints and integrality restrictions on some or all of the variables (Nemhauser & Wolsey, 1999). In particular models consist of linear, integer (representing binary choice) 0–1 variables. Therefore, the optimization models presented in this chapter are defined as mixed integer programming problems.

An efficient solution to the multi-criteria portfolio optimization problem can be found by applying the lexicographic and weighting approach (Ehrgott, 2000; Sawik, 2007a, 2007b, 2008c; Steuer, 1986; Wiecek, 2007).

The lexicographic optimization generates efficient solutions that can be found by sequential optimization with elimination of the dominating functions. The weighted objective functions also generate various efficient solutions. It provides a complete parametrization of the efficient set for multi-criteria mixed integer programs.

Table 1. Notations.

Indices	
i	Historical time period, $i \in I = \{1, \ldots, m\}$
j	Asset/stock, $j \in J = \{1, \ldots, n\}$
Input parameters	
p_i	Probability assigned to the occurrence of past realization i
r_{ij}	Observed return of jth stock in ith time period
r^{Min}	Minimum return observed in the market
r^{VaR}	Return Value-at-Risk
α^{VaR}	Input parameter in problem (M2), (M4) – probability that return of investment is not less than r^{VaR}
r^*	Minimum expected return that the decision maker is prepared to accept
υ	Accepted number of stocks in optimal portfolio
β_1,β_2,β_3	Weights in the objective function
Decision variables	
λ_j	Amount of capital invested in jth stock
y_i	1 if return of portfolio in ith time period is over threshold r^{VaR}, 0 otherwise
α^{VaR}	Decision variable in problem (M1), (M3) – probability that return of investment is not less than r^{VaR}
z_j	1 if capital is invested in jth stock, 0 otherwise

The first optimization model presented in this chapter is the triple objective portfolio optimization model with weighting approach, that is, the minimization of risk probability of portfolio loss versus maximization of expected portfolio return versus maximization of amount of capital to be invested in portfolio (Model **M1**).

The second and following models are based on lexicographic and weighting approach. The optimality criterion is to maximize expected portfolio return (model **M2**) subject to various constraints with two decision variables λ_j and y_i (Table 1).

The obtained solution value (model **M2**) is used as model parameter for the bi-objective function (model **M3**), the minimization of risk probability of portfolio loss versus the maximization of amount of capital to be invested in portfolio subject to selected constraints with three decision variables λ_j, y_i, and α^{VaR} (Table 1).

The solution values from problem (models **M2** and **M3**) are used as input parameters for optimization problem (model **M4**), that is, the minimization of number of stocks in optimal portfolio subject to selected constraints with three decision variables λ_j, y_i, and z_j (Table 1).

The decision variable α^{VaR} in models **M1** and **M3** is an input parameter for models **M2** and **M4**.

PROBLEM FORMULATION

Suppose that n assets are available in the financial market with historical quotations in m time periods. Let r_{ij} be the random variable representing the future return of jth asset in ith time instant. The portfolio optimization problem with *VaR* constraint is formulated as the classic Markowitz approach, but with *VaR* instead of variance as risk measure.

The decision maker fixes two parameters, the probability α^{VaR} and lower bound for successful returns – any investment whose *VaR* is less than r^{VaR} will be not acceptable (Benati & Rizzi, 2007; Sawik, 2007a, 2007b, 2008a, 2008b, 2008c).

Let r^{Min} be the minimum return that can be observed in the market, for example, the biggest possible loss of money invested in portfolio (Benati & Rizzi, 2007). In the worst case it is the whole amount of capital, so for instance r^{Min} can be equal -100%.

OPTIMIZATION MODELS

The multi-objective portfolio optimization model with lexicographic and weighting approach is presented below.

The portfolio optimization model (**M1**) with weighting approach deals with three following objective functions, that is, the minimization of risk probability of portfolio loss versus maximization of expected portfolio return versus maximization of amount of capital to be invested in portfolio.

The primary objective is to maximize expected portfolio return (model **M2**), then the minimization of risk probability of portfolio loss versus the maximization of amount of capital to be invested in portfolio is considered (model **M3**), and finally, the minimization of number of stocks in optimal portfolio (model **M4**) is achieved.

The lexicographic and weighting multi-objective portfolio optimization model with *VaR* are NP-hard problems even when future returns are described by discrete uniform distributions (Nemhauser & Wolsey, 1999).

In the approach proposed in this chapter, the portfolio optimization problem is formulated as a triple objective mixed integer program, which allows commercially available software (e.g., AMPL/CPLEX; Fourer, Gay, & Kerninghan, 1993) to be applied for solving medium size, yet practical instances.

Model M1. Maximize

$$-\beta_1 \alpha^{VaR} + \beta_2 \sum_{i=1}^{m} p_i \left(\sum_{j=1}^{n} \lambda_j r_{ij} \right) + \beta_3 \sum_{j=1}^{n} \lambda_j \tag{1}$$

subject to

$$y_i \leq \frac{\sum_{j=1}^{n} \lambda_j r_{ij} - r^{\text{Min}}}{r^{VaR} - r^{\text{Min}}}, \quad i \in I \tag{2}$$

$$\sum_{i=1}^{m} p_i (1 - y_i) \leq \alpha^{VaR} \tag{3}$$

$$\sum_{j=1}^{n} \lambda_j \leq 1 \tag{4}$$

$$\lambda_j \geq 0; \quad j \in J : \sum_{i=1}^{m} p_i r_{ij} > 0 \tag{5}$$

$$0 \leq \alpha^{VaR} \leq 1 \tag{6}$$

$$y_i \in \{0, 1\}, \quad i \in I \tag{7}$$

Model M2. Minimize

$$\sum_{i=1}^{m} p_i \left(\sum_{j=1}^{n} \lambda_j r_{ij} \right) \tag{8}$$

subject to Eqs. (2–5) and (7).

Model M3. Minimize

$$-\beta_1 \alpha^{VaR} + \beta_2 \sum_{j=1}^{n} \lambda_j \tag{9}$$

subject to Eqs. (2–7) and

$$\sum_{i=1}^{m} p_i \left(\sum_{j=1}^{n} \lambda_j r_{ij} \right) \geq r^* \tag{10}$$

Model M4. Minimize

$$\sum_{j=1}^{n} z_j \tag{11}$$

subject to Eqs. (2–5), (7), (10), and

$$\sum_{j=1}^{n} z_j \leq v \tag{12}$$

$$\lambda_j \leq z_j, \quad j \in J \tag{13}$$

$$z_j \in \{0, 1\}, \quad j \in J \tag{14}$$

Variables λ_j are percentage of wealth that is allocated to asset j. Constraints (2) and (3) prevent the choice of portfolios whose VaR is below the fixed threshold. Every time expected portfolio return is below r^{VaR}, then decision variable y_i must be equal to 0 and $1 - y_i = 1$ in constraint (3). Therefore, all probabilities of events i whose returns are below the VaR threshold were summed up. If the result is greater than α^{VaR}, then the portfolio is not feasible (Benati & Rizzi, 2007).

The combination of continuous variables λ_j and α^{VaR} and of binary variable y_i leads this mixed integer programming problem. If the number of historical observations m is bounded by a constant, there are 2^m ways of fixing the decision variables y_i (Nemhauser & Wolsey, 1999).

Constraint (4) requires that not more than one unit of wealth must be allocated on different assets. Constraint (5) defines continuous variable λ_j – amount of capital invested in jth stock. This formula prevents short-selling and in addition includes quasi-cutting constraint for the elimination of stocks with non-positive expected return. Constraints (6), (7), and (14) define decision variables. Constraint (10) imposes the minimum portfolio expected return r^* that the decision maker is prepared to accept. Constraint (12) ensures that the number of stocks in optimal portfolio must be less than or equal to accepted number of assets in selected portfolio. Finally, constraint (13) is responsible for dependency between variables λ_j and z_j.

COMPUTATIONAL RESULTS

In this section, numerical examples and some computational results are presented to illustrate possible applications of the proposed formulations of a lexicographic and weighting approach by mixed integer program and to compare the results. The examples are modeled on a real data from the Warsaw Stock Exchange. In computational experiments, the historical time period is 1,758 days (seven years), 500 and 100 days with 139 stocks considered. Computational time range is from a few seconds to minutes or even hours. The computational experiments have been performed using AMPL with CPLEX 9.1 on a PC Compaq Presario 1830 with Pentium III; RAM 512 MB and on a PC Compaq Presario 3000 Pentium III, RAM 512 MB.

Table 2 presents the influence of different parameters on CPU run time. The computational experiments for $\alpha^{VaR} = 0.5$, that is, when median is considered as a risk or an efficiency criterion, have indicated that only one stock is selected, so that only trivial solutions are obtained (Sawik, 2007a, 2007b, 2008a, 2008b, 2008c).

Table 3 presents solution results for the weighting approach (model **M1**) with 100 historical quotations.

Column ''number of assets'' defines amount of stocks in optimal solutions.

In the tables, column ''MIP simplex iteration'' shows the number of mixed integer programming simplex iterations until the solution is presented.

Column ''B-&-B nodes'' shows the number of searched nodes in the branch and bound tree until the solution presented.

Column ''GAP'' shows percentage difference between obtained solution and the best LP-relaxation based bound calculated by the CPLEX solver.

Table 4 presents solution results for the weighting approach (model **M1**) with 500 historical quotations.

Table 2. Problem Parameters versus Central Processing Unit Run Time.

α^{VaR} increases	CPU decreases
r^{VaR} increases	CPU increases
m increases	CPU increases

Note: α^{VaR} input parameter for optimization problems (model **M2**) and (model **M4**).

Table 3. The Solution Results for the Weighting Approach (Model **M1**) with 100 Historical Quotations.

β_1	β_2	β_3	α^{VaR}	Portfolio Return	Amount of Capital	Number of Assets	MIP Simplex Iteration	B-&-B Nodes	CPU/GAP (%)
0.80	0.10	0.10	0.000	0.658086	1	7	1061	101	4.11
0.10	0.80	0.10	0.280	1.738270	1	1	56	0	0.17
0.10	0.10	0.80	0.270	1.731680	1	2	112	17	0.82
0.70	0.15	0.15	0.040	0.861917	1	13	25788	3910	65.30
0.15	0.70	0.15	0.280	1.738270	1	1	56	0	1.59
0.15	0.15	0.70	0.270	1.731680	1	2	112	17	0.77
0.60	0.20	0.20	0.230	1.633050	1	8	4741	1008	19.17
0.20	0.60	0.20	0.280	1.738270	1	1	56	0	0.16
0.20	0.20	0.60	0.270	1.731680	1	2	112	17	0.71
0.50	0.25	0.25	0.260	1.713360	1	4	712	124	4.12
0.25	0.50	0.25	0.280	1.738270	1	1	63	0	0.22
0.25	0.25	0.50	0.270	1.731680	1	2	112	17	0.71
0.40	0.30	0.30	0.270	1.731680	1	2	188	37	1.92
0.30	0.40	0.30	0.270	1.731680	1	2	76	2	0.50
0.30	0.30	0.40	0.270	1.731680	1	2	112	17	0.88

Note: CPU seconds for proving optimality on a PC Compaq Presario 1830 Pentium III, RAM 512MB/CPLEX 9.1.

Table 4. The Solution Results for the Weighting Approach (Model **M1**) with 500 Historical Quotations.

β_1	β_2	β_3	α^{VaR}	Portfolio Return	Amount of Capital	Number of Assets	MIP Simplex Iteration	B-&-B Nodes	CPU/GAP
0.80	0.10	0.10	0.018	0.473987	1	35	190977472	7709994	71271.21
0.10	0.80	0.10	0.360	1.295300	1	1	361	1	0.51%
0.10	0.10	0.80	0.338	1.280820	1	6	7896177	1660831	4280.88
0.70	0.15	0.15	0.042	0.600772	1	34	62253914	5882993	22595.40
0.15	0.70	0.15	0.360	1.295300	1	1	766	21	1.00%
0.15	0.15	0.70	0.340	1.281930	1	5	11217491	2391321	5431.07
0.60	0.20	0.20	0.062	0.666516	1	34	30208779	5402504	12736.50
0.20	0.60	0.20	0.360	1.295300	1	1	3419	901	4.72%
0.20	0.20	0.60	0.338	1.280660	1	6	6192835	1402627	3520.78
0.50	0.25	0.25	0.274	1.186270	1	10	23069048	5320236	39033.00
0.25	0.50	0.25	0.358	1.294360	1	2	114568	61391	164.66
0.25	0.25	0.50	0.334	1.277550	1	6	41459235	10512354	26839.90
0.40	0.30	0.30	0.332	1.273370	1	9	29281271	5992131	13337.90
0.30	0.40	0.30	0.340	1.282980	1	4	2047901	560025	1731.51
0.30	0.30	0.40	0.336	1.278810	1	6	49893067	11875407	30246.80
0.34	0.33	0.33	0.336	1.278810	1	6	36991485	7984601	20446.20

Note: CPU seconds for proving optimality on a PC Compaq Presario 3000 Pentium III, RAM 512MB/CPLEX 9.1.

Table 5 shows the solution results for maximization of expected portfolio return (model **M2**) for 1758.

Table 6 presents the results for the maximization of expected portfolio return (model **M2**) for 500 historical time periods.

Table 5. The Results for the Maximization of Expected Portfolio Return (Model **M2**) for 1758 Quotations.

α^{VaR}	r^{VaR}	Portfolio Return	Amount of Capital	Number of Assets	MIP Simplex Iteration	B-&-B Nodes	GAP	CPU
0.10	−2.00	0.406521	1.0000	16	22203	2601	10.24%	3671.41
0.15	−2.00	0.450744	1.0000	8	43873	9401	0.66%	3599.88
0.15	−1.00	0.357077	1.0000	33	17534	1101	7.14%	3600.14
0.15	−0.50	0.223021	0.8436	61	105900	2601	99.43%	32182.40
0.50	−0.25	0.109703	0.3421	28	3813	100	319.88%	1176.50

Note: CPU seconds for proving optimality on PC Compaq Presario 1830 with Pentium III, RAM 512 MB/CPLEX 9.1.

Table 6. The Results for the Maximization of Expected Portfolio Return (Model **M2**) for 500 Quotations.

α^{VaR}	r^{VaR}	Portfolio Return	Amount of Capital	Number of Assets	MIP Simplex Iteration	B-&-B Nodes	GAP/ CPU
0.01	−10.00	1.221675	1.0000	3	47	6	3.35
0.01	−5.00	0.937749	1.0000	10	1044	91	19.72
0.01	−4.00	0.849007	1.0000	10	2588	192	31.42
0.01	−3.00	0.743942	1.0000	13	12139	512	84.64
0.01	−2.00	0.629044	1.0000	17	33359	1105	242.11
0.01	−1.50	0.532917	1.0000	23	111944	2498	544.31
0.01	−1.00	0.396546	1.0000	30	297576	5462	1689.62
0.01	−0.50	0.198811	0.5314	31	826053	11333	5295.86
0.01	−0.25	0.099405	0.2657	31	671318	8527	4690.36
0.05	−10.00	1.295303	1.0000	1	34	0	0.88
0.05	−5.00	1.225334	1.0000	5	273	22	7.47
0.05	−4.00	1.141190	1.0000	9	11445	2450	161.97
0.05	−3.00	1.020270	1.0000	13	334933	33001	2.84%
0.05	−2.00	0.883416	1.0000	19	86761	3401	16.13%
0.10	−3.00	1.160440	1.0000	10	53836	6801	2.87%
0.15	−2.00	1.118940	1.0000	14	32074	3790	7.73%
0.20	−2.00	1.221670	1.0000	7	39648	7501	1.87%
0.25	−1.50	1.234330	1.0000	8	38871	8801	1.56%

Table 7. The Solution Results for the Weighting Approach (Model **M3**) with 1758 Historical Quotations.

β_1	β_2	r^{VaR}	r^*	α^{VaR}	Portfolio Return	Amount of Capital	Number of Assets	MIP Simplex Iteration	CPU
0.9	0.1	−2.00	0.40	0.111490	0.400819	1	4	222	4.78
0.5	0.5	−2.00	0.40	0.112059	0.400921	1	4	222	4.95
0.1	0.9	−2.00	0.40	0.114334	0.400459	1	4	224	4.73
0.9	0.1	−2.00	0.45	0.151308	0.450515	1	3	294	5.33
0.5	0.5	−2.00	0.45	0.150171	0.450000	1	3	294	5.06
0.1	0.9	−2.00	0.45	0.150171	0.450000	1	3	293	4.73
0.9	0.1	−1.00	0.34	0.134243	0.340825	1	18	615	10.10
0.5	0.5	−1.00	0.34	0.135381	0.340647	1	18	671	10.00
0.1	0.9	−1.00	0.34	0.133675	0.340722	1	18	721	12.03
0.9	0.1	−0.50	0.20	0.110353	0.202772	1	56	1040	30.81
0.5	0.5	−0.50	0.20	0.111490	0.204490	1	56	1123	34.82
0.1	0.9	−0.50	0.20	0.110353	0.202382	1	56	1185	37.07
0.9	0.1	−0.25	0.10	0.192833	0.141738	1	66	1748	49.60
0.5	0.5	−0.25	0.10	0.191126	0.142552	1	68	1733	49.98
0.1	0.9	−0.25	0.10	0.191695	0.142652	1	68	1559	57.72

Table 8. The Solution Results for the Weighting Approach (Model **M3**) with 500 Historical Quotations.

r^{VaR}	r^*	α^{VaR}	Portfolio Return	Amount of Capital	Number of Assets	MIP Simplex Iteration	CPU
−10.00	1.22168	0.010	1.221670	1	3	14	2.31
−5.00	0.94774	0.016	0.937749	1	6	124	1.04
−4.00	0.84900	0.014	0.849007	1	9	180	1.32
−3.00	0.74394	0.016	0.752234	1	12	213	3.35
−2.00	0.62900	0.016	0.629000	1	12	254	3.79
−1.50	0.53300	0.020	0.533000	1	21	198	3.74
−1.00	0.39560	0.022	0.421036	1	21	295	3.46
−0.50	0.19900	0.050	0.219878	1	30	392	3.46
−0.25	0.09940	0.104	0.232379	1	41	499	4.78
−10.00	1.29530	0.022	1.295300	1	1	19	3.13
−5.00	1.22530	0.060	1.226390	1	2	39	1.53
−4.00	1.14000	0.060	1.140000	1	8	47	3.02
−3.00	1.02000	0.058	1.020000	1	8	54	0.88
−2.00	0.88000	0.070	0.880000	1	12	101	1.15
−3.00	1.16044	0.124	1.160440	1	7	86	0.88
−2.00	1.11894	0.160	1.122190	1	8	106	1.10
−2.00	1.22167	0.212	1.221670	1	5	119	0.88
−1.50	1.23433	0.264	1.234330	1	5	146	0.99

Note: $\beta_1 = 0.5$, $\beta_2 = 0.5$; Objective function: α^{VaR} and portfolio return.

Table 9. Number of Assets in Optimal Portfolio for Lexicographic Approach (Model **M4**) with 1758 Historical Quotations.

α^{VaR}	r^{VaR}	r^*	Number of Assets (υ)	$\sum_{j=1}^{n} z_j$
0.10	−2.00	0.40	16	4
0.15	−2.00	0.45	8	3
0.15	−1.00	0.34	33	18
0.15	−0.50	0.20	61	56
0.50	0.25	0.10	68	68

Table 10. Number of Assets in Optimal Portfolio for Lexicographic Approach (Model **M4**) with 500 Historical Quotations.

α^{VaR}	r^{VaR}	r^*	Number of Assets (υ)	$\sum_{j=1}^{n} z_j$
0.01	−10.00	1.2217	3	3
0.01	−5.00	0.9377	10	6
0.01	−4.00	0.8490	10	9
0.01	−3.00	0.7439	13	12
0.01	−2.00	0.6290	17	12
0.01	−1.50	0.5330	23	21
0.01	−1.00	0.3956	30	21
0.01	−0.50	0.1990	31	30
0.01	−0.25	0.0994	31	31
0.05	−10.00	1.2953	1	1
0.05	−5.00	1.2253	5	2
0.05	−4.00	1.1400	9	8
0.05	−3.00	1.0200	13	8
0.05	−2.00	0.8800	19	12
0.10	−3.00	1.1604	10	7
0.15	−2.00	1.1189	14	8
0.20	−2.00	1.2217	7	5
0.25	−1.50	1.2343	8	5

Table 7 presents the solution results for the weighting approach (model **M3**) with 1758 historical quotations.

Table 8 shows the solution results for the weighting approach (model **M3**) with 500 historical quotations.

Table 9 presents the number of assets in optimal portfolio for lexicographic approach (model **M4**) with 1758 historical quotations.

Table 11. Examples of CPU Time for Computational Experiments for Optimal Portfolio for Lexicographic Approach (Model **M4**) with 1758 Historical Quotations.

α^{VaR}	r^{VaR}	r^*	MIP Simplex Iteration	B-&-B Nodes	CPU
0.01	−10	1.2217	27	10	110.02
0.01	−5	0.9377	610	192	880.79
0.01	−4	0.8490	3217	1024	5045.62
0.01	−3	0.7439	7779	2300	10272.80
0.01	−2	0.6290	19980	4014	21677.90

Table 10 shows the number of assets in optimal portfolio for lexicographic approach (model **M4**) with 500 historical quotations.

Table 11 presents examples of CPU time for computational experiments for optimal portfolio for lexicographic approach (model **M4**) with 1,758 historical quotations.

The computational time for the optimization model with objective function (model **M4**) requires several CPU minutes for finding the first feasible solution.

The total computational time ranges from a few seconds to minutes or even hours depending on the number of historical quotations in the optimization problem.

CONCLUSIONS

This chapter shows the lexicographic and weighting approach and the corresponding mixed integer programming formulations for the multi-criteria portfolio optimization problem.

The optimal solution value for objective function, the maximization of amount of capital to be invested in portfolio, has proven to be constant for the considered instances of the problem. The computational time range to find proven optimal solution for model **M4** requires minutes, hours, or even days.

The computational experiments modeled on a real data from the Warsaw Stock Exchange have indicated that the approach is capable of finding proven optimal solutions for medium size problems in a reasonable computation time using commercially available software for mixed integer programming.

ACKNOWLEDGMENT

This work has been supported by Polish Ministry of Science & Higher Education (MNISW) grant for PhD Research #N N519 405934.

REFERENCES

Benati, S., & Rizzi, R. (2007). A mixed integer linear programming formulation of the optimal mean/value-at-risk portfolio problem. *European Journal of Operational Research*, *176*, 423–434.

Ehrgott, M. (2000). *Multicriteria optimization* (2nd ed.). Berlin: Springer.

Esch, L., Kieffer, R., Lopez, T., Berb, C., Damel, P., Debay, M., & Hannosset, J.-F. (2005). *Asset and risk management. Risk oriented finance*. John Wiley & Sons.

Fourer, R., Gay, D., & Kerninghan, B. (1993). AMPL-A modeling programming language for mathematical programming. The Scientific Press Series, USA.

Markowitz, H. M. (1952). Portfolio selection. *Journal of Finance*, *7*, 77–91.

Markowitz, H. M. (1959). *Portfolio selection: Efficient diversification of investments*. New York: Wiley.

Nemhauser, G. L., & Wolsey, L. A. (1999). *Integer and combinatorial optimization*. New York, Toronto, Canada: Wiley.

Ogryczak, W. (2000). Multiple criteria linear programming model for portfolio selection. *Annals of Operations Research*, *97*, 143–162.

Sarykalin, S., Serraino, G., Uryasev, S. (2008) Value-at-Risk vs. Conditional Value-at-risk in risk management and optimization. In: Z.-L. Chen, S. Raghavan, & P. Gray (Eds), *Tutorials in operations research*, INFORMS Annual Meeting, October 12–15, 2008.

Sawik, B. (2007a). A multi-objective portfolio optimization by mixed integer programming. *The 20th anniversary conference of the European chapter on combinatorial optimization (ECCO XX)*, Limassol, Cyprus, May 24–26, 2007.

Sawik, B. (2007b). Weighting vs. lexicographic approach to multi-objective portfolio optimization by mixed integer programming. In: E. Kochan (Ed.), *Problems of mechanical engineering and robotics, monographic of faculty of mechanical engineering and robotics, AGH* (Vol. 36). Poland: Kraków.

Sawik, B. (2008a). A triple-objective portfolio optimization by mixed integer linear programming. In: H. Howaniec & W. Waszkielewicz (Eds), *Company management-financial aspects, information-communication and operational.* Monographic of ATH University of Bielsko-Biala, Poland.

Sawik, B. (2008b) A weighted-sum mixed integer program for multi-criteria portfolio optimization. *The 21st conference of the European chapter on combinatorial optimization (ECCO XXI), Dubrovnik*, Croatia, May 29–31, 2008.

Sawik, B. (2008c). A three stage lexicographic approach for multi-criteria portfolio optimization by mixed integer programming. *Przegląd Elektrotechniczny*, *84*(9), 108–112.

Sharpe, W. F. (1971). A linear programming approximation for the general portfolio analysis problem. *Journal of Financial and Quantitative Analysis*, *6*, 1263–1275.

Speranza, M. G. (1993). Linear programming models for portfolio optimization. *Finance*, *14*, 107–123.

Steuer, R. E. (1986). *Multiple criteria optimization: Theory, computation and application*. New York: Wiley.

Wiecek, M. M. (2007). Advances in cone-based preference modeling for decision making with multiple criteria. *Decision Making in Manufacturing and Services, AGH*, *1*(1–2), 153–173.

EXTREME MEAN-VARIANCE SOLUTIONS: ESTIMATION ERROR VERSUS MODELING ERROR[☆]

Robert R. Grauer

ABSTRACT

Without short-sales constraints, mean-variance (MV) and power-utility portfolios generated from historical data are often characterized by extreme expected returns, standard deviations, and weights. The result is usually attributed to estimation error. I argue that modeling error, that is, modeling the portfolio problem with just a budget constraint, plays a more fundamental role in determining the extreme solutions and that a more complete analysis of MV problems should include realistic constraints, estimates of the means based on predictive variables, and specific values of investors' risk tolerances. Empirical evidence shows that investors who utilize MV analysis without imposing short-sales constraints, without employing estimates of the means based on predictive variables, and without specifying their risk tolerance miss out on remarkably remunerative investment opportunities.

[☆]The paper was presented at the Ninth Conference on Pacific Business, Economics and Finance at Rutgers University, the Northern Finance Association meetings in Halifax, and the Pacific Northwest Finance Conference in Seattle.

Financial Modeling Applications and Data Envelopment Applications
Applications of Management Science, Volume 13, 19–51

ISSN: 0276-8976/doi:10.1108/S0276-8976(2009)0000013004

1. INTRODUCTION

Without short-sales constraints, mean-variance (MV) and power-utility portfolios generated from historical data are often characterized by extreme expected returns, standard deviations, and weights. The results are often attributed to estimation error – particularly in the means. Research on portfolio selection has focused on three ways to generate more meaningful solutions.

A first approach constrains the portfolio weights. Frost and Savarino's (1988) simulation analysis shows that imposing upper and lower bounds on MV portfolio weights reduces estimation bias and improves performance. With short-sales constraints imposed, Grauer and Hakansson (1986, 1987, 1993, 1995) find power-utility and MV asset allocation strategies often generate economically and statistically significant returns in a variety of out-of-sample settings. But, even with short-sales constraints imposed, Black and Litterman (1992) decry the seemingly "unreasonable and unbalanced" nature of the composition of MV portfolios and argue that investors should "shrink" their estimates of the means to equilibrium means.[1]

A second approach employs estimates of the means based on statistical, financial, or forecasting models. Jobson, Korkie, and Ratti (1979), Jobson and Korkie (1980, 1981), and Jorion (1985, 1986, 1991) present simulation and out-of-sample evidence that suggests statistical estimators (known as Stein, Bayes–Stein (BS), Empirical Bayes, or shrinkage estimators) and financial (capital asset pricing model (CAPM)) estimators can improve MV portfolios' investment performance substantially. Grauer and Hakansson (1995), employing a power-utility optimizer and short-sales constraints, confirm that Stein estimators outperform the sample estimator in an industry rotation setting, but the gains are not as great as those reported by others. In a global setting, just the opposite is true. The sample estimator outperforms the Stein estimators. In all cases, the CAPM estimator exhibits the worst performance – just the opposite of what Jorion (1991) finds in an industry setting using MV analysis that allows short sales.

A third approach concentrates on the global minimum-variance portfolio of risky assets. Eun and Resnick (1994), Jaganathan and Ma (2003), and Jorion (1985, 1986, 1991) demonstrate that in the absence of short-sales constraints, the minimum-variance portfolio almost invariably outperforms the tangency portfolio out-of-sample when judged by the Sharpe ratio. Moreover, Chan, Karceski, and Lakonishok (1999), among others, focus exclusively on the global minimum-variance portfolio.[2] But, given the returns Grauer (2008a) generates from a power-utility model, which takes into account short-sales constraints, estimates of the means based on predictive variables, and investor risk tolerance, arguments that these variables should be ignored in a MV framework are open to question.[3]

Although estimation error undoubtedly plays a part in determining the extreme MV solutions, I argue that modeling error, that is, modeling the portfolio problem with just a budget constraint, plays a more essential role and that a more complete analysis of MV problems should include realistic constraints, estimates of the means based on predictive variables, and specific investor risk tolerance values. The argument is straightforward. Two basic characteristics of MV problems have fundamental implications for active asset allocation. First, given a covariance matrix and the weights in the "market" (or any benchmark) portfolio, Best and Grauer (1985) show that it is easy to construct sets of means that insure the market portfolio is the tangency portfolio. Specifically, let Σ be the covariance matrix and $\mathbf{x}_m$ be the vector of "market" portfolio weights, then the $(\boldsymbol{\Sigma}, \mathbf{x}_m)$ compatible means are $\boldsymbol{\mu} = \theta_r \boldsymbol{\iota} + \theta_m \boldsymbol{\Sigma} \mathbf{x}_m$, where $\theta_r = \lambda(T_m)/T_m = r$ and $\theta_m = (\mu_m - r)/\sigma_m^2 = 1/T_m$ are scalar parameters, μ_m and σ_m^2 are the expected return and the variance of the market portfolio, respectively, $\boldsymbol{\Sigma}\mathbf{x}_m$ is an n-vector whose jth element, $\mathrm{cov}(r_j, r_m)$, is the covariance of the return on security j with the return on the market portfolio, T_m is the risk tolerance parameter of a "representative investor" who neither borrows nor lends, and r is the riskless interest rate. Given the covariance matrix and the $(\boldsymbol{\Sigma}, \mathbf{x}_m)$ compatible means, a representative MV investor will hold the tangency portfolio $\mathbf{x}_m$.[4] Second, Best and Grauer (1991, 1992) document the extreme sensitivity of portfolio weights to perturbations in the $(\boldsymbol{\Sigma}, \mathbf{x}_m)$ compatible means.[5]

In turn, these basic characteristics of MV problems have fundamental implications for active asset allocation. First, "representative" investors who employ these equilibrium means will by default adopt the *passive* policy of holding the market (i.e., the tangency) portfolio. Second, *active* managers who hope to generate abnormal returns must employ means that have forecasting ability. But, given the acute sensitivity of the portfolio weights to small perturbations in the equilibrium means, they must also impose short-sales constraints to avoid extreme positions – positions that would be disallowed in practice.[6] Furthermore, in the long run, the extreme positions will almost certainly result in large out-of-sample losses. Therein lays the dilemma. In a CAPM setting, employing the MV problem with just a budget constraint is genius. The math is both elegant and simple. In equilibrium, no one attempts to short sell. Everyone holds the market portfolio combined with either borrowing or lending. In a real-world investment setting, employing the MV problem with just a budget constraint is modeling error.

A more complete analysis of MV problems should include realistic constraints, estimates of the means based on predictive variables, and specific values of investors' risk tolerances.[7]

To make the case, I compare the policies and performance of the global minimum-variance portfolio, the tangency portfolio, and six MV efficient portfolios generated with and without short-sales constraints when the means are estimated in four different ways. The analysis employs quarterly decision horizons in an industry rotation setting that spans the 1934–1999 period. When short sales are permitted, many of the solutions are nothing short of bizarre. But the results are easily explained in terms of the well-known efficient set mathematics.[8] If there is no riskfree asset, the minimum-variance frontier of risky assets is a hyperbola in mean–standard deviation space. If there is a riskless asset, the position of the tangency portfolio depends on the relative values of the riskless rate and the mean of the global minimum-variance portfolio. If the riskless rate is less than (greater than) the mean of the global minimum-variance portfolio, the tangency portfolio will be MV efficient (MV *inefficient*). If the riskless rate is just slightly less than the mean of the global minimum-variance portfolio, the expected return, standard deviation, and weights of the tangency portfolio will take on extreme values. When short sales are precluded, the solutions are much more realistic. More important, many of the conclusions drawn from the analysis when short sales are permitted are completely reversed when short sales are precluded. And, much of the richness of the MV model is lost if we ignore investor risk tolerance and simply focus on the minimum-variance and tangency portfolios.

The chapter proceeds as follows. Section 2 formulates MV problems, with short-sales permitted and with short-sales precluded, and explains the method employed to make the model operational. The MV problems are expressed in terms of investor risk tolerance. This formulation relates directly to the efficient set mathematics and has a natural link to power utility. Section 3 describes the data. Section 4 presents four alternative ways of estimating the means. Sections 5 and 6 contain the empirical results. Section 7 includes a summary and conclusions.

2. THE MV MODEL

Following Markowitz (1959), Sharpe (1970), and Best and Grauer (1990), the MV problem is

$$\max(T\mu_p - 1/2\sigma_p^2) \tag{1}$$

subject to a budget constraint (and perhaps other constraints), where μ_p is one plus the rate of return on the portfolio, σ_p^2 is the variance of the return on the portfolio, and T is a scalar parameter. We may think of the T as a parametric quadratic programming parameter, where the minimum-variance frontier is traced out as T varies from $-\infty$ to $+\infty$. We may also think of T as a (fixed positive) risk tolerance parameter.[9] In this case, Eq. (1) is an investor's "MV utility function," where the larger the T is, the more tolerant the investor is to risk. Alternatively, the MV utility function may be viewed as an approximation to a more basic von Neumann–Morgenstern utility function. More specifically, it may be viewed as an approximation to the isoelastic family of utility functions $U(w) = (1/\gamma)w^\gamma$, for some $\gamma < 1$, where $U(w)$ denotes the utility of wealth and γ is a risk aversion parameter. Using a Taylor-series approximation to expected utility for short holding periods, it can be shown that T is equal to the reciprocal of the Pratt–Arrow measure of relative risk aversion (RRA), where RRA is equal to $-wU''(w)/U'(W)$ evaluated at a zero rate of return. Thus, T is a risk tolerance parameter. For a power function with risk aversion parameter γ, RRA is equal to $(1-\gamma)$, and the MV approximation to it sets $T = 1/(1-\gamma)$. Under certain conditions, the approximation holds exactly in continuous time (Merton, 1973). Furthermore, Grauer and Hakansson (1993) show that the MV approximation works well for power utility investors with quarterly decision horizons when short sales are precluded.

At each time (quarter) t, the simplest MV problem that includes a riskfree asset is

$$\max T(\boldsymbol{\mu}_t'\mathbf{x}_t + r_t x_{rt}) - 1/2\mathbf{x}_t'\boldsymbol{\Sigma}_t\mathbf{x}_t + \lambda_t(1 - \boldsymbol{\iota}'\mathbf{x}_t - x_{rt}) \tag{2}$$

where $\boldsymbol{\mu}$, $\mathbf{x}$, and $\boldsymbol{\iota}$ are n-vectors containing one plus the expected rates of return, portfolio weights, and ones, respectively; r is unity plus the riskfree rate of return; x_r is the fraction of wealth invested in the riskless asset, with $x_r > 0$ denoting lending, and $x_r < 0$ denoting borrowing; Σ is an $n \times n$ positive-definite covariance matrix of asset returns; λ is the Lagrange multiplier associated with the budget constraint $\boldsymbol{\iota}'\mathbf{x} + x_r = 1$; and by definition, $\mu_p = \boldsymbol{\mu}'\mathbf{x} + rx_r$ and $\sigma_p^2 = \mathbf{x}'\boldsymbol{\Sigma}\mathbf{x}$. For ease of exposition, this problem is referred to as the MV problem "with short sales permitted."

A more realistic formulation of the MV problem is

$$\max \quad T\left(\boldsymbol{\mu}_t'\mathbf{x}_t + r_{Lt}x_{Lt} + r_{Bt}^d x_{Bt}\right) - 1/2\mathbf{x}_t'\boldsymbol{\Sigma}_t\mathbf{x}_t \tag{3}$$

subject to

$$x_{it} \geq 0, \text{ all } i, \; x_{Lt} \geq 0, \quad x_{Bt} \leq 0, \tag{4}$$

$$\sum_i x_{it} + x_{Lt} + x_{Bt} = 1 \tag{5}$$

$$\sum_i m_{it} x_{it} \leq 1 \tag{6}$$

where x_{Lt} is the amount lent in period t, r_{Lt} is one plus the riskfree lending rate in period t, x_{Bt} is the amount borrowed in period t, r^d_{Bt} is one plus the riskfree borrowing rate at the time of the decision at the beginning of period t, and m_{it} is the initial margin requirement for asset category i in period t expressed as a fraction. Constraint (4) rules out short sales and ensures that lending (borrowing) is a positive (negative) fraction of capital. Constraint (5) is the budget constraint. Constraint (6) serves to limit borrowing (when desired) to the maximum permissible under the margin requirements that apply to the various asset categories. This problem is referred to as the MV problem "with short sales precluded," even though it also includes margin constraints and borrowing and lending at different rates.

At the beginning of quarter t, the MV problem (Eqs. (3)–(6)) for that quarter uses the following inputs: the (observable) riskfree return for quarter t, the (observable) call money rate $+1\%$ at the beginning of quarter t, and the (observable) realized returns for the risky asset categories for the previous k quarters. The observable returns (together with the returns on the market) are used to estimate $\boldsymbol{\mu}_t$ and $\boldsymbol{\Sigma}_t$. The different ways of estimating $\boldsymbol{\mu}_t$ are discussed in Section 4. With these inputs in place, the portfolio weights $\mathbf{x}_t$ for the various asset categories and the proportion of assets borrowed or lent are calculated by solving the system (Eqs. (3)–(6)) through quadratic programming methods. For the single-constraint MV problem (Eq. (2)), with riskfree lending and borrowing at the same rate, the closed-form solutions are

$$\mathbf{x}_t = T\,\boldsymbol{\Sigma}_t^{-1}\left(\boldsymbol{\mu}_t - r_t\boldsymbol{\iota}\right) \quad \text{and} \quad x_{rt} = 1 - T(r_t c_t - a_t) \tag{7}$$

where $a = \boldsymbol{\iota}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ and $c = \boldsymbol{\iota}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\iota}$ are two of the three well-known efficient set constants. At the end of quarter t, the realized returns on the risky assets are observed, along with the realized borrowing rate r^r_{Bt} (which may differ from the decision borrowing rate r^d_{Bt}). Then, using the weights selected at the beginning of the quarter, the realized return on the portfolio chosen for quarter t is recorded. The cycle is repeated in all subsequent quarters. All reported returns are gross of transaction costs and taxes and assume that the investor in question had no influence on prices. There are several reasons for this approach. First, as in previous studies, I wish to keep the complications to a minimum. Second, the return series used as inputs and for comparisons also exclude transaction costs (for reinvestment of interest and dividends) and taxes. Third, many investors are tax-exempt and various techniques are available for keeping transaction costs low. Finally, since the proper treatment of these items is nontrivial, they are better left to a later study.

3. THE DATA

The data used to estimate mean vectors and covariance matrices and to calculate the realized returns on the portfolios come from several sources. The returns for the value-weighted industry groups are constructed from the returns on individual New York Stock Exchange firms contained in the Center for Research in Security Prices' (CRSP) monthly returns database. The firms are combined into twelve industry groups on the basis of the first two digits of the firms' SIC codes. (The work by Grauer, Hakansson, and Shen (1990) contains a detailed description of the industry data.) The riskfree asset is assumed to be 90-day U.S. Treasury bills maturing at the end of the quarter. The *Survey of Current Business* and the *Wall Street Journal* are the sources. The borrowing rate is assumed to be the call money rate +1% for decision purposes (but not for rate of return calculations). The applicable beginning of period decision rate, r^d_{Bt}, is viewed as persisting throughout the period and thus as riskfree. For 1934–1976, the call money rates are obtained from the *Survey of Current Business*. For later periods, the *Wall Street Journal* is the source. Finally, margin requirements for stocks are obtained from the *Federal Reserve Bulletin*.[10]

4. FOUR ESTIMATORS OF THE MEANS

Historical estimators of the means were the first to be employed in the literature. The n-vector of historic means at the beginning of period t is denoted as

$$\boldsymbol{\mu}_{Ht} = (\bar{r}_{1t}, \ldots, \bar{r}_{nt})' \tag{8}$$

where $\bar{r}_{it} = 1/k(\sum_{\tau=t-k}^{t-1} r_{i\tau})$ and $k = 32$. The choice of 32 quarters for the "moving" window follows the discrete-time power-utility literature. (In the MV literature, the moving window is usually 60 months in length.) This historical approach estimates the means one at a time, relying exclusively on information contained in each of the time series.

Stein's (1955) suggestion that the efficiency of the estimate of the means could be improved by pooling the information across series led to an alternative set of so-called shrinkage estimators that shrink the historical means to some grand mean. A classic example is the James–Stein estimator (Efron % Morris, 1973, 1975, 1977). It was first employed in the portfolio selection literature by Jobson et al. (1979). This chapter, however, focuses on Jorion's (1985, 1986, 1991) BS estimator

$$\boldsymbol{\mu}_{BSt} = (1 - w_t)\boldsymbol{\mu}_{Ht} + w_t \bar{r}_{Gt}\boldsymbol{\iota} \tag{9}$$

where $w_t = \lambda_t/(\lambda_t + k)$ is the shrinking factor, $\lambda_t = (n+2)/((\boldsymbol{\mu}_{Ht} - \bar{r}_{Gt}\boldsymbol{\iota})' \widehat{\boldsymbol{\Sigma}}_t^{-1}(\boldsymbol{\mu}_{Ht} - \bar{r}_{Gt}\boldsymbol{\iota}))$, n is the number of risky assets, $\widehat{\boldsymbol{\Sigma}}_t$ is the sample covariance matrix calculated from the k periods in the estimation period, and $\bar{r}_{Gt} = \boldsymbol{\iota}'\widehat{\boldsymbol{\Sigma}}_t^{-1}\boldsymbol{\mu}_{Ht}/(\boldsymbol{\iota}'\widehat{\boldsymbol{\Sigma}}_t^{-1}\boldsymbol{\iota})$ is the grand mean. In this case, the grand mean is the mean of the global minimum-variance portfolio generated from the historical data.

A third estimator of the means is based on the Sharpe (1964) and Lintner (1965) CAPM. The CAPM predicts that all assets plot on the security market line (SML)

$$\boldsymbol{\mu} = r\,\boldsymbol{\iota} + \left(\mu_m - r\right)\boldsymbol{\beta}$$

where μ_m is the expected return on the market portfolio, the jth element of the n-vector $\boldsymbol{\beta} \equiv \boldsymbol{\Sigma}\mathbf{x}_m/\sigma_m^2$ is the covariance of the return on asset j with the return on the market portfolio divided by the variance of the return on the market portfolio. There are a number of ways in which one might estimate CAPM means. The most popular is

$$\boldsymbol{\mu}_{CAPMt} = r_{Lt}\boldsymbol{\iota} + (\bar{r}_{mt} - \bar{r}_{Lt})\hat{\boldsymbol{\beta}}_t \tag{10}$$

where $\bar{r}_{mt} = 1/k(\sum_{\tau=t-k}^{t-1} r_{m\tau})$, $\bar{r}_{Lt} = 1/k(\sum_{\tau=t-k}^{t-1} r_{L\tau})$, and $\bar{r}_{mt} - \bar{r}_{Lt}$ is an estimate of the expected excess return on the "market" portfolio. At each time t, $\hat{\boldsymbol{\beta}}_t$ is estimated from the market model regressions

$$r_{i\tau} = \alpha_{it} + \beta_{it} r_{m\tau} + e_{i\tau}, \quad \text{for all } i \text{ and } \tau$$

in the estimation period from $t-k$ to $t-1$. Following convention, the CRSP value-weighted index is employed as the proxy for the market portfolio.

A fourth estimator of the mean vector uses dividend yields and riskfree interest rates to forecast the means. At each time t, the regression

$$r_{i\tau} = a_{0i} + a_{1i} dy_{\tau-1} + a_{2i} r_{L\tau} + e_{i\tau}, \quad \text{for all } i \text{ and } \tau \tag{11}$$

is run in the $t-k$ to $t-1$ estimation period, where $dy_{\tau-1}$ is the annual CRSP value-weighted dividend yield lagged one month and $r_{L\tau}$ is the quarterly riskfree (T-bill) rate. Both independent variables are "de-meaned." Hence, a_{0i} is the historic average rate of return on asset i. The traditional one-period ahead forecast of the mean of industry i is $\bar{r}_{DRit} = \hat{a}_{0i} + \hat{a}_{1i} dy_{t-1} + \hat{a}_{2i} r_{Lt}$, where $\hat{a}_{0i}, \hat{a}_{2i}$, and $\hat{a}_{2i}$ are the estimated coefficients, and dy_{t-1} and r_{Lt} are observable at the beginning of period $t+1$. That is, the quarterly variable dy_{t-1} is lagged one month, and there is no need to lag r_{Lt} as it is observable at the beginning of the quarter. The vector of dividend-yield riskfree-rate (DR) estimators is

$$\boldsymbol{\mu}_{DRt} = (\bar{r}_{DR1t}, \ldots, \bar{r}_{DRnt})' \tag{12}$$

However, this forecast is extremely variable. Therefore, in the spirit of the Stein estimators, I "shrink" the DR means to historic means. Assuming that the DR and the historic means are equally likely, the vector of dividend-yield riskfree-rate-historic (DRH) mean estimators is[11]

$$\boldsymbol{\mu}_{DRHt} = (\boldsymbol{\mu}_{DRt} + \boldsymbol{\mu}_{Ht})/2 \tag{13}$$

5. RESULTS FOR THE 1934–1999 PERIOD

Space limitations preclude reporting a complete set of results or even commenting on all the results reported in the tables. The ex post performance of selected benchmark portfolios, and the ex ante and ex post performance, accumulated wealth levels, and the weights of the MV portfolios are reported in turn. The reported results are generated from a 32-quarter moving window but are robust to 96-month, 20-quarter, and 60-month estimation periods.

Table 1 reports the ex post performance of selected benchmark portfolios: the twelve value-weighted industries, T-bills, the unlevered "market" portfolio constructed by value weighting the twelve industries, and levered market portfolios that either borrow or lend. The table shows that \$1 invested in Treasury bills at the beginning of 1934 grows to \$14 at the end of 1999; \$1 invested in the market grows to \$1,882; and \$1 invested in the market, levered to a minimum of 200% or the applicable maximum percentage dictated by the initial margin constraints set by the Federal Reserve, grows to \$11,720. The ex post Sharpe ratios of the market portfolio and the levered market portfolios that lend are 0.27, and the Sharpe ratios of the levered market portfolios that borrow are 0.25. The differences in the borrowing and lending portfolios' Sharpe ratios reflect the fact that the borrowing rate exceeds the lending rate.

Table 1. Ex Post Performance of Selected Benchmark Portfolios.

	Ex Post Mean (%)	Ex Post Standard Deviation (%)	Sharpe Ratio	Wealth ($)	Compound Return (%)
Panel A: Industry portfolios					
Petroleum	3.5	9.5	0.26	2,578	3.02
Finance and real estate	3.5	9.9	0.25	2,578	3.02
Consumer durables	3.7	11.0	0.24	3,084	3.09
Basic industries	3.3	9.0	0.25	1,800	2.88
Food and tobacco	3.3	8.0	0.28	2,211	2.96
Construction	3.5	11.7	0.21	1,710	2.86
Capital goods	3.4	9.8	0.25	2,047	2.93
Transportation	3.0	11.7	0.17	426	2.32
Utilities	3.0	7.2	0.27	1,134	2.70
Textiles and trade	3.5	10.5	0.23	2,047	2.93
Services	4.2	14.4	0.22	4,087	3.20
Leisure	3.9	13.1	0.22	3,084	3.09
Treasury bills	1.0	0.8	0.00	14	1.01
Panel B: Levered value-weighted market (VW) portfolios					
50% VW, 50% Lending	2.1	4.2	0.27	209	2.04
100% VW	3.2	8.3	0.27	1,882	2.90
150% VW, 50% Borrowing	4.1	12.3	0.25	5,693	3.33
200% VW, 100% Borrowing	4.8	15.5	0.25	11,720	3.61

Notes: The investment universe consists of twelve value-weighted industry groups in the 1934–1999 period. The Sharpe ratio is defined as mean excess return divided by standard deviation of excess return. Excess return is measured relative to the lending rate. Wealth is the cumulative wealth at the end of 1999 arising from an investment of $1 at the beginning of 1934. The value-weighted market portfolio (VW) value weights the twelve value-weighted industry portfolios. At each point in time, the 50% (100%) borrowing portfolios can invest the minimum of 150% (200%) of wealth or the applicable maximum percentage dictated by the initial margin constraints set by the Federal Reserve. These portfolios borrow at a rate that exceeds the lending rate. Returns are measured in percent per quarter.

Table 2 describes the ex ante and ex post performance of six MV efficient portfolios, the global minimum-variance portfolios, and the tangency portfolios generated from four sets of means with short sales permitted. To be consistent with the literature that permits short sales, it is assumed that the borrowing rate equals the lending rate and that there are no margin constraints. The results are bizarre. Ex ante many of the tangency portfolios are *inefficient*! Ex post the tangency portfolios – except those based on the CAPM means – together with most of the more risk tolerant MV efficient portfolios bankrupt! Surprisingly, one would never know this from an examination of ex post Sharpe ratios. In Table 2, many non-bankrupt and

Table 2. The Performance of Six Mean-Variance Efficient Portfolios, the Minimum-Variance Portfolio, and the Tangency Portfolios Generated from Four Sets of Means with Short Sales Permitted.

Portfolio	Average Ex Ante Mean (%)	Average Ex Ante Standard Deviation (%)	Average Ex Ante Sharpe Ratio	Ex Post Mean (%)	Ex Post Standard Deviation (%)	Ex Post Sharpe Ratio	Wealth ($)	Compound Return (%)
Panel A: Historic means								
$\gamma = -50$	2.7	1.7	0.88	1.6	3.7	0.16	53	1.51
$\gamma = -10$	8.7	8.0	0.88	3.6	16.9	0.16	320	2.21
$\gamma = -5$	15.2	14.7	0.88	5.8	31.0	0.16	0	−0.99
$\gamma = -1$	43.5	44.0	0.88	15.4	93.1	0.16	0	–
$\gamma = 0$	86.0	87.9	0.88	29.8	186.2	0.16	0	–
$\gamma = 0.5$	170.9	175.8	0.88	58.6	372.4	0.16	0	–
Minimum variance	2.0	4.4	0.28	2.2	7.4	0.16	159	1.94
Tangency	1.0	55.5	0.51	51.5	617.0	0.08	0	–
Panel B: Bayes–Stein means								
$\gamma = -50$	1.8	1.1	0.54	1.5	2.7	0.19	50	1.49
$\gamma = -10$	4.5	4.9	0.54	3.4	12.5	0.19	1,060	2.67
$\gamma = -5$	7.3	9.0	0.54	5.4	22.8	0.19	2,133	2.95
$\gamma = -1$	20.0	27.1	0.54	14.0	68.5	0.19	0	–
$\gamma = 0$	39.0	54.2	0.54	27.1	137.0	0.19	0	–
$\gamma = 0.5$	76.9	108.3	0.54	53.1	274.0	0.19	0	–
Minimum variance	2.0	4.4	0.28	2.2	7.4	0.16	159	1.94
Tangency	2.9	19.4	0.35	15.8	174.2	0.09	0	–
Panel C: CAPM means								
$\gamma = -50$	1.2	0.5	0.28	1.2	0.8	0.24	23	1.20
$\gamma = -10$	2.1	2.5	0.28	1.9	3.7	0.24	117	1.82
$\gamma = -5$	3.0	4.6	0.28	2.6	6.7	0.24	509	2.39
$\gamma = -1$	6.8	13.8	0.28	5.8	20.1	0.24	2,537	3.01
$\gamma = 0$	12.6	27.6	0.28	10.6	40.2	0.24	0	–
$\gamma = 0.5$	24.2	55.2	0.28	20.1	80.4	0.24	0	–
Minimum variance	1.6	4.4	0.14	2.2	7.4	0.16	159	1.94
Tangency	3.1	9.9	0.27	3.4	8.4	0.28	2,442	3.00
Panel D: Dividend-yield riskfree-rate means								
$\gamma = -50$	3.4	2.1	1.06	1.7	4.3	0.17	73	1.64
$\gamma = -10$	12.2	9.7	1.06	4.3	19.8	0.17	384	2.28
$\gamma = -5$	21.5	17.7	1.06	7.1	36.3	0.17	0	–
$\gamma = -1$	62.4	53.1	1.06	19.3	108.8	0.17	0	–

Table 2. (*Continued*)

Portfolio	Average Ex Ante Mean (%)	Average Ex Ante Standard Deviation (%)	Average Ex Ante Sharpe Ratio	Ex Post Mean (%)	Ex Post Standard Deviation (%)	Ex Post Sharpe Ratio	Wealth ($)	Compound Return (%)
$\gamma = 0$	123.8	106.2	1.06	37.5	217.5	0.17	0	–
$\gamma = 0.5$	246.5	212.4	1.06	74.1	435.1	0.17	0	–
Minimum variance	1.9	4.4	0.24	2.2	7.4	0.16	159	1.94
Tangency	52.0	95.6	0.44	−35.3	792.9	−0.05	0	–

Notes: A mean-variance efficient portfolio is defined in terms of an approximation to a power-utility function $u(w) = 1/\gamma w^{\gamma}$. The corresponding risk tolerance parameter employed in the mean-variance optimizer is $T = 1/(1-\gamma)$. The investment universe consists of twelve value-weighted industries in the 1934–1999 period. Quarterly portfolio revision with a 32-quarter estimation period is employed. Borrowing and lending take place at the riskfree lending rate. Returns are measured in percent per quarter. The ex post standard deviation used in constructing the ex post Sharpe ratio is measured in units of excess return. Wealth is the cumulative wealth at the end of 1999 arising from an investment of $1 at the beginning of 1934.

bankrupt strategies share the same Sharpe ratios. Moreover, some bankrupt portfolios have greater Sharpe ratios than non-bankrupt portfolios. Consider, for example, two portfolios with $\gamma = 0.5$ and 0 generated from CAPM means. The Sharpe ratios of these bankrupt MV efficient portfolios when short sales are permitted (see Table 2) are greater than the Sharpe ratios of the corresponding MV efficient portfolios when short sales are precluded (Table 3). Yet, over the 1934–1999 period, a $1 investment in the latter two portfolios grows to $3,523 and $6,341, respectively.

For a given set of means, the average values of the ex ante Sharpe ratios of the tangency portfolio and the six MV efficient portfolios differ. The difference in the Sharpe ratios occurs because the MV efficient portfolios – characterized by positive risk tolerance parameters – are efficient *even when the tangency portfolio is not*.[12] Although it is not reported explicitly, the four tangency portfolios (which are listed in the same order as in Table 2, e.g., the tangency portfolio generated from historic means first, the tangency portfolio generated from BS means second, the tangency portfolio generated from CAPM means third, and the tangency portfolio generated from DR means fourth) are ex ante *inefficient* 62, 62, 15, and 78 times, respectively.

The zero wealth levels reported in Table 2 indicate that a number of portfolios lost over 100 percent in at least one of the 264 quarters in the

Table 3. The Performance of Six Mean-Variance Efficient Portfolios, the Minimum-Variance Portfolio, and the Tangency Portfolios Generated from Four Sets of Means with Short Sales Precluded.

Portfolio	Average Ex Ante Mean (%)	Average Ex Ante Standard Deviation (%)	Average Ex Ante Sharpe Ratio	Ex Post Mean (%)	Ex Post Standard Deviation (%)	Ex Post Sharpe Ratio	Wealth ($)	Compound Return (%)
Panel A: Historic means								
$\gamma = -50$	1.5	0.8	0.42	1.4	1.4	0.26	36	1.37
$\gamma = -10$	2.9	3.6	0.42	2.3	5.2	0.26	315	2.20
$\gamma = -5$	3.9	5.9	0.40	3.0	7.6	0.26	1,024	2.66
$\gamma = -1$	5.9	11.7	0.38	4.4	12.6	0.26	10,234	3.56
$\gamma = 0$	6.8	15.7	0.36	4.9	14.9	0.26	21,354	3.85
$\gamma = 0.5$	7.6	19.8	0.35	5.8	18.1	0.26	69,015	4.31
Minimum variance	2.8	7.2	0.32	3.1	7.2	0.28	1,508	2.81
Tangency	4.5	11.1	0.42	3.4	10.4	0.23	1,701	2.86
Panel B: Bayes–Stein means								
$\gamma = -50$	1.3	0.6	0.30	1.3	1.2	0.24	30	1.30
$\gamma = -10$	2.1	2.4	0.29	2.1	4.2	0.25	177	1.98
$\gamma = -5$	2.6	3.8	0.28	2.4	5.7	0.25	376	2.27
$\gamma = -1$	3.6	7.8	0.25	3.6	9.9	0.26	2,900	3.07
$\gamma = 0$	4.0	10.4	0.24	4.3	11.9	0.27	10,570	3.57
$\gamma = 0.5$	4.3	13.4	0.22	4.7	13.8	0.26	15,409	3.72
Minimum variance	2.3	7.2	0.25	3.1	7.2	0.28	1,508	2.81
Tangency	2.8	9.3	0.28	3.3	9.4	0.24	1,626	2.84
Panel C: CAPM means								
$\gamma = -50$	1.2	0.5	0.27	1.2	0.8	0.24	23	1.20
$\gamma = -10$	2.1	2.5	0.27	1.9	3.6	0.24	117	1.82
$\gamma = -5$	2.7	4.2	0.27	2.3	5.4	0.23	263	2.13
$\gamma = -1$	4.0	9.1	0.26	3.5	9.8	0.25	2,552	3.02
$\gamma = 0$	4.7	12.8	0.24	3.9	12.5	0.23	3,523	3.14
$\gamma = 0.5$	5.3	17.4	0.23	4.6	15.3	0.23	6,341	3.37
Minimum variance	2.4	7.2	0.23	3.1	7.2	0.28	1,508	2.81
Tangency	3.1	10.0	0.27	3.3	8.4	0.28	2,447	3.00
Panel D: Dividend-yield riskfree-rate means								
$\gamma = -50$	1.5	0.8	0.43	1.4	1.4	0.27	38	1.39
$\gamma = -10$	3.1	3.7	0.42	2.6	5.6	0.27	521	2.40
$\gamma = -5$	4.2	6.1	0.41	3.3	8.2	0.28	2,228	2.96
$\gamma = -1$	6.4	11.9	0.39	5.2	13.7	0.31	63,988	4.28

Table 3. (*Continued*)

Portfolio	Average Ex Ante Mean (%)	Average Ex Ante Standard Deviation (%)	Average Ex Ante Sharpe Ratio	Ex Post Mean (%)	Ex Post Standard Deviation (%)	Ex Post Sharpe Ratio	Wealth ($)	Compound Return (%)
$\gamma = 0$	7.5	15.9	0.37	6.2	17.0	0.30	186,769	4.70
$\gamma = 0.5$	8.5	20.3	0.36	7.2	20.9	0.30	437,116	5.04
Minimum variance	2.5	7.2	0.26	3.1	7.2	0.28	1,508	2.81
Tangency	4.6	11.0	0.41	3.6	10.2	0.26	3,302	3.12

Notes: A mean-variance efficient portfolio is defined in terms of an approximation to a power-utility function $u(w) = 1/\gamma w^{\gamma}$. The corresponding risk tolerance parameter employed in the mean-variance optimizer is $T = 1/(1-\gamma)$. The investment universe consists of twelve value-weighted industries in the 1934–1999 period. Quarterly portfolio revision with a 32-quarter estimation period is employed. The borrowing rate exceeds the lending. Returns are measured in percent per quarter. The ex post standard deviation used in constructing the ex post Sharpe ratio is measured in units of excess return. Wealth is the cumulative wealth at the end of 1999 arising from an investment of $1 at the beginning of 1934.

1934–1999 period. Again, although it is not reported explicitly, the four tangency portfolios in Table 2 lost more than 100 percent 4, 2, 0, and 10 times. The more risk-tolerant MV-efficient portfolios bankrupted considerably more often. For example, the MV approximation to the 0.5 power lost more than 100 percent 68, 49, 10, and 73 times, respectively.

Continuing on with the results in Table 2, note that the global minimum-variance portfolios are estimated without reference to the means. Hence, the ex post performance is identical in all four panels, as is the ex ante standard deviation. Of course, the ex ante means and Sharpe ratios differ (but not dramatically) with different estimates of the means. Over the 1934–1999 period, $1 invested in the minimum-variance portfolio grows to $159 and its ex post Sharpe ratio is 0.16, neither of which compares favorably with a passive investment in the market portfolio. By way of contrast, the tangency portfolio differs for each set of means. The average of the ex ante Sharpe ratios ranges from 0.27 for the BS means to 0.51 for the historic means. Consistent with Jorion (1991), the ex post performance of the tangency portfolios estimated from the CAPM means exceeds that of the global minimum-variance portfolio, even though the ex post performance of the other tangency portfolios is abysmal.[13]

Table 3 reports dramatically different results when short sales are precluded, the borrowing rate exceeds the lending rate, and margin constraints are imposed. Three results stand out. First, although the minimum-variance portfolio does not keep pace with the market portfolio in terms of accumulated wealth, its performance improves considerably. Over the 1934–1999 period, \$1 invested in the minimum-variance portfolio grows to \$1,508 versus \$159 when short sales are permitted. In addition, it's ex post Sharpe ratio is slightly greater than that of the market portfolio. Second, the performance of the tangency portfolios not generated from CAPM means improve dramatically. Rather than bankrupting, a \$1 investment in the tangency portfolio grows from a minimum of \$1,626, with BS historic means, to a maximum of \$3,302, with DR means. Third, the improvement in the performance of the tangency portfolios pales in comparison to the improvement in performance exhibited by the more risk-tolerant MV-efficient portfolios generated from DR estimates of the means. With short sales permitted, the MV-efficient portfolio that approximates the 0.5 power bankrupts. With short sales precluded, a \$1 investment in this portfolio at the beginning of 1934 grows to over \$437,000 at the end of 1999. The means that generate portfolios exhibiting the worst performance with short sales permitted generate portfolios exhibiting the best performance with short sales precluded!

Tables 2 and 3 focus on average values of the variables. But averages only tell part of the story. Therefore, Table 4 reports the minimum and maximum values for the returns (together with the average values for comparative purposes) of the minimum-variance portfolio and the tangency portfolios generated from the historic, CAPM, and DR means. With short sales permitted, the minimum and maximum values of the ex ante means and ex post returns of the tangency portfolio generated from historic and DR means are virtually unbelievable. The minimum (maximum) ex ante mean is –1,301 (12,580) percent per quarter. (Naturally, the ex ante standard deviations share the same unbelievable characteristics as the ex ante means.) Not surprisingly, the minimum (maximum) ex post return is equally extreme: −12,543 (1,767) percent per quarter. With short sales precluded, the return characteristics of the portfolios are credible.

Tables 5 and 6 examine the portfolio weights rather than returns or wealth levels. Table 5 reports the results for the minimum-variance and tangency portfolios when short sales are permitted. In light of the ex ante means and standard deviations of the tangency portfolios generated from the historic and DR means, it should come as no surprise that the minimum and maximum portfolio weights take on near surrealistic values. With historic means, the minimum (maximum) weight in any industry is −272 (288). With DR means, the minimum (maximum) weight is even more extreme taking on

Table 4. Descriptive Statistics for the Returns of Minimum-Variance and Tangency Portfolios Generated from Three Sets of Means with Short Sales Permitted and Short Sales Precluded.

	Minimum-Variance Portfolio				Tangency Portfolio			
	Ex Ante Mean (%)	Ex Ante Standard Deviation (%)	Ex Ante Sharpe Ratio	Ex Post Return (%)	Ex Ante Mean (%)	Ex Ante Standard Deviation (%)	Ex Ante Sharpe Ratio	Ex Post Return (%)
Panel A: Historic means, short sales permitted								
Minimum	−0.8	2.0	−0.61	−29.3	−2020.7	2.6	−1.16	−1,006.2
Average	2.0	4.4	0.28	2.2	1.0	55.5	0.51	51.5
Maximum	5.0	6.6	1.28	18.4	503.1	3773.8	1.68	9,770.6
Panel B: Historic means, short sales precluded								
Minimum	−0.1	2.8	−0.19	−22.1	0.0	3.1	−0.22	−30.9
Average	2.8	7.2	0.32	3.1	4.5	11.1	0.42	3.4
Maximum	4.9	17.7	1.04	37.1	9.0	41.7	1.06	58.2
Panel C: CAPM means, short sales permitted								
Minimum	0.0	2.0	−0.05	−29.3	0.4	4.9	−0.09	−24.8
Average	1.6	4.4	0.14	2.2	3.1	9.9	0.27	3.4
Maximum	3.9	6.6	0.42	18.4	6.1	26.4	0.89	35.1
Panel D: CAPM means, short sales precluded								
Minimum	0.2	2.8	−0.09	−22.1	0.4	5.0	−0.09	−24.8
Average	2.4	7.2	0.23	3.1	3.1	10.0	0.27	3.3
Maximum	4.3	17.7	0.69	37.1	6.2	26.6	0.89	35.8
Panel E: Dividend-yield riskfree-rate means, short sales permitted								
Minimum	−2.8	2.0	−1.02	−29.3	−1,301.1	2.7	−1.90	−12,543.4
Average	1.9	4.4	0.24	2.2	52.0	95.6	0.44	−35.3
Maximum	5.5	6.6	1.46	18.4	12,580.5	12,278.6	1.79	1,767.0
Panel F: Dividend-yield riskfree-rate means, short sales precluded								
Minimum	−4.1	2.8	−0.83	−22.1	−8.2	3.1	−0.75	−32.7
Average	2.5	7.2	0.26	3.1	4.6	11.0	0.41	3.6
Maximum	6.7	17.7	1.29	37.1	15.8	41.2	1.41	46.5

Notes: The investment universe consists of twelve value-weighted industry groups in the 1934–1999 period. Quarterly portfolio revision with a 32-quarter estimation period is employed. When short sales are permitted, borrowing and lending take place at the riskfree lending rate. When short sales are precluded, the borrowing rate exceeds the lending rate. The tangency portfolio is constructed using the lending rate. Returns are measured in percent per quarter. The Sharpe ratio is defined as the mean excess return divided by the standard deviation of return. If the optimal policy is to lend everything, that is, if there is no feasible tangency portfolio, then the ex ante Sharpe ratio is set to zero.

Table 5. Descriptive Statistics for the Weights of Minimum-Variance and Tangency Portfolios Generated from Three Sets of Means with Short Sales Permitted.

Industry	Tangency Portfolio Historic Means			Tangency Portfolio CAPM Means		
	Minimum	Average	Maximum	Minimum	Average	Maximum
Petroleum	−162	0.72	188	0.02	0.14	0.21
Finance and real estate	−102	−0.59	204	−0.07	0.06	0.28
Consumer durables	−87	2.00	81	0.03	0.15	0.31
Basic industries	−191	−0.21	93	0.00	0.17	0.25
Food and tobacco	−101	1.64	49	−0.06	0.07	0.16
Construction	−32	−0.81	35	−0.06	0.02	0.18
Capital goods	−213	−0.74	233	0.00	0.11	0.24
Transportation	−104	−0.21	66	−0.08	0.03	0.12
Utilities	−214	−0.77	23	0.05	0.15	0.29
Textiles and trade	−67	−0.17	100	−0.07	0.06	0.14
Services	−106	1.06	288	−0.08	0.01	0.18
Leisure	−272	−0.92	130	−0.12	0.03	0.19
	Tangency Portfolio Dividend-Yield Riskfree-Rate Means			**Minimum-Variance Portfolio**		
	Minimum	Average	Maximum	Minimum	Average	Maximum
Petroleum	−72	4.85	1,115	−0.54	0.27	0.97
Finance and real estate	−407	−1.19	101	−2.02	−0.68	−0.07
Consumer durables	−1,923	−3.89	558	−0.70	0.25	1.35
Basic industries	−3,504	−14.05	83	−0.86	0.18	1.14
Food and tobacco	−57	9.27	2,041	−0.72	0.52	1.70
Construction	−1,276	−4.13	138	−0.86	−0.15	0.78
Capital goods	−338	−0.85	158	−0.72	0.03	1.17
Transportation	−35	8.12	2,159	−0.47	0.03	0.80
Utilities	−483	−2.69	40	−0.14	0.77	1.52
Textiles and trade	−383	1.69	858	−0.96	−0.04	0.70
Services	−23	7.78	1,874	−0.39	0.03	0.58
Leisure	−610	−3.91	15	−1.06	−0.19	0.61

Notes: The investment universe consists of twelve value-weighted industry groups in the 1934–1999 period. Quarterly portfolio revision with a 32-quarter estimation period is employed. When short sales are permitted, borrowing and lending take place at the riskfree lending rate. The portfolio weights are numbers that sum to one, not percentages that sum to 100.

Table 6. Descriptive Statistics for the Weights of Minimum-Variance and Tangency Portfolios Generated from Three Sets of Means with Short Sales Precluded.

Industry	Tangency Portfolio Historic Means				Tangency Portfolio CAPM Means			
	Minimum	Average	Maximum	Positive	Minimum	Average	Maximum	Positive
Petroleum	0.01	0.38	1.00	151	0.02	0.14	0.20	264
Finance and real estate	0.06	0.28	1.00	16	0.00	0.07	0.28	229
Consumer durables	0.00	0.22	1.00	45	0.03	0.14	0.28	264
Basic industries	0.00	0.29	0.97	51	0.00	0.17	0.26	264
Food and tobacco	0.01	0.55	1.00	87	0.00	0.08	0.17	245
Construction	0.01	0.38	1.00	29	0.00	0.04	0.15	166
Capital goods	0.01	0.48	1.00	58	0.00	0.11	0.24	264
Transportation	0.02	0.15	0.42	14	0.00	0.04	0.11	234
Utilities	0.02	0.45	1.00	82	0.05	0.15	0.25	264
Textiles and trade	0.00	0.36	0.81	39	0.00	0.06	0.11	240
Services	0.00	0.36	1.00	86	0.00	0.03	0.16	172
Leisure	0.03	0.25	0.68	28	0.00	0.03	0.15	221

	Tangency Portfolio Dividend-Yield Riskfree-Rate Means				Minimum-Variance Portfolio			
	Minimum	Average	Maximum	Positive	Minimum	Average	Maximum	Positive
Petroleum	0.00	0.41	1.00	127	0.00	0.21	0.54	187
Finance and real estate	0.00	0.50	1.00	25	0.00	0.00	0.00	0

Consumer durables	0.01	0.29	1.00	40	0.00	0.07	0.22	25
Basic industries	0.01	0.35	1.00	34	0.02	0.19	0.41	59
Food and tobacco	0.02	0.52	1.00	69	0.01	0.45	1.00	135
Construction	0.05	0.47	1.00	28	0.00	0.06	0.17	16
Capital goods	0.00	0.42	1.00	60	0.01	0.07	0.13	10
Transportation	0.01	0.20	0.81	16	0.06	0.07	0.08	2
Utilities	0.00	0.48	1.00	90	0.00	0.60	0.93	226
Textiles and trade	0.03	0.46	1.00	35	0.00	0.14	0.31	27
Services	0.02	0.43	1.00	66	0.01	0.14	0.33	59
Leisure	0.04	0.32	0.79	36	0.06	0.12	0.16	10

Notes: The investment universe consists of twelve value-weighted industry groups in the 1934–1999 period. Quarterly portfolio revision with a 32-quarter estimation period is employed. The borrowing rate exceeds the lending rate. The tangency portfolio is constructed using the lending rate. The minimum, average, and maximum weights are taken with respect to positive weights only, that is, zero weights are not included in the calculation. Zero values are due to rounding. Positive denotes the number of times (out of 264 quarters) that an industry is held with a positive weight. The weights are reported as decimal fractions that sum to one.

a value of −3,504 (2,159). (The portfolio weights sum to one. They are not percentages that sum to 100.) Of course, no real world investor could possibly establish such extreme long and short positions. On the contrary, although the tangency portfolios generated from CAPM means contain negative weights, the minimum and maximum weights are much more sensible, ranging from −0.12 to 0.31. The weights in the minimum-variance portfolio are not quite as well behaved. The weights range from −2.02 to 1.70, the maximum investment in finance and real estate is −0.07, and the average weights for finance and real estate, construction, textiles and trade, and leisure are negative.

Table 6 reports the corresponding results with short sales precluded. In this case, the weights in the tangency portfolios are much more reasonable. However, the portfolios are characterized by plunging behavior with most industries reaching minimum and maximum weights of zero and one, respectively. To provide more insight into this plunging behavior, Table 6 also records the number of times (out of 264 quarters) each industry is held with positive weights. The portfolios generated from the CAPM means hold most of the industries most of the time. Perhaps surprisingly, the minimum-variance portfolio eschews finance and real estate completely and only holds transportation twice. Finally, although it is not reported in Table 6, another way to see the differences in the weights is in terms of the number of risky assets held in the tangency and minimum-variance portfolios at a point in time when short sales are precluded. The global minimum-variance portfolio and the tangency portfolios generated from historic and DR means never hold *more* than six industries at any point in time. The tangency portfolios generated from CAPM means never hold *less* than seven.

I do not report the results of statistical tests in this chapter for two reasons. First, the power-utility portfolios with short sales precluded reported in Grauer (2008a) are similar to the MV-efficient portfolios with short sales precluded reported here. That is, the power-utility portfolios generated from DR estimators outperform those based on historic, BS, and CAPM estimators of the means. With some exceptions, commonly accepted statistical measures of investment performance support the rankings.

Second, performance measures analyze arithmetic rather than compound returns. This could, and in fact does, lead to anomalous results when portfolios bankrupt. As noted in Table 2, many non-bankrupt and bankrupt portfolios share the same Sharpe ratios. I chose to include the results of statistical tests of the bankrupt strategies from the work by Grauer (2008b), as the results are even more anomalous than I expected. That work examines whether popular performance measures can distinguish between two extremes: perfect-foresight and bankrupt asset allocation strategies. Bankruptcy is the ultimate investment risk. Perfect-foresight strategies yield returns beyond anyone's wildest dreams. Yet, the unconditional and conditional Jensen and Fama–French alphas and the Grinblatt–Titman portfolio change measures of the bankrupt MV-efficient portfolios in Table 2, panel D, are greater than the alphas and portfolio change measures of the corresponding MV-efficient portfolios in Table 3, panel D – one of which grew from \$1 to over \$437,000 in the 1934–1999 period. Worse, the unconditional and conditional Jensen and Fama–French alphas of two bankrupt MV portfolios and the Grinblatt–Titman portfolio change measure of three bankrupt MV portfolios are greater than the alphas and portfolio change measures of *all* the perfect-foresight portfolios!

6. RESULTS FOR THE FOURTH QUARTER OF 1984

Section 5 examined the effect of different constraint sets and means employing time series data. This section examines what happens at a point in time. The fourth quarter of 1984 is chosen to illustrate the most extreme portfolio characteristics and outcomes; the sensitivity of portfolio weights, ex ante returns, and realized outcomes to different sets of means; and the importance of specifying investor risk tolerance. Table 7 reports the ex ante and ex post returns of the six MV-efficient portfolios, the minimum-variance portfolio, and the tangency portfolio with and without short-sales constraints in the fourth quarter of 1984. The results for the portfolios generated from DR means, historic means, and CAPM means are shown in panels A, B, and C, respectively. The corresponding minimum-variance frontiers are presented in Figs. 1, 2, and 3. Each of the figures shows the asset means and standard deviations and four minimum-variance frontiers – with and without riskfree borrowing and lending and with and without short-sales constraints. When short sales are permitted, the minimum-variance frontier that allows riskfree borrowing and lending at the same rate is shown as a dotted line.

Panel A shows that when short sales are permitted, the ex ante expected return of the tangency portfolio generated from DR means is in excess of 12,500 percent per quarter – its weights range from $-3{,}504$ to 2,158 times wealth – and ironically its realized return exceeds $-12{,}500$ percent! The characteristics of the six MV-efficient portfolios, while still excessive in some cases, are much more sensible. This is easily explained in terms of the efficient set mathematics. It is well known (e.g., see Eq. (7)) that an investor holds the unlevered tangency portfolio if his implied risk tolerance parameter is

Table 7. Returns of Six Constant Risk Tolerance, the Minimum-Variance and Tangency Portfolios Generated from Three Sets of Means with Short Sales Permitted and Precluded in the Fourth Quarter of 1984.

Portfolio	Short Sales Permitted				Short Sales Precluded			
	Ex Ante Mean (%)	Ex Ante Standard Deviation (%)	Ex Ante Sharpe Ratio	Realized Return (%)	Ex Ante Mean (%)	Ex Ante Standard Deviation (%)	Ex Ante Sharpe Ratio	Realized Return (%)
Panel A: Dividend-yield riskfree-rate means								
$\gamma = -50$	4.7	2.0	1.02	0.6	2.7	0.1	0.05	2.7
$\gamma = -10$	12.2	9.3	1.02	−6.9	2.7	0.4	0.05	2.8
$\gamma = -5$	20.1	17.1	1.02	−14.8	2.7	0.8	0.05	2.9
$\gamma = -1$	55.1	51.2	1.02	−49.7	2.8	2.4	0.05	3.3
$\gamma = 0$	107.6	102.4	1.02	−102.0	2.9	4.8	0.05	4.0
$\gamma = 0.5$	212.5	204.9	1.02	−206.7	3.0	6.6	0.05	4.5
Minimum variance	2.7	3.4	0.00	7.6	2.7	5.3	0.01	5.5
Tangency	12,580.5	12,278.6	1.02	−12,543.4	3.0	6.6	0.05	4.5
Panel B: Historic means								
$\gamma = -50$	4.7	2.0	1.01	−0.5	2.8	0.6	0.30	2.5
$\gamma = -10$	11.9	9.2	1.01	−11.9	3.5	2.7	0.30	2.0
$\gamma = -5$	19.7	16.8	1.01	−24.0	4.2	5.0	0.30	1.4
$\gamma = -1$	53.7	50.5	1.01	−77.2	6.5	13.0	0.30	−0.6
$\gamma = 0$	104.7	101.0	1.01	−157.1	9.1	23.0	0.28	−3.3
$\gamma = 0.5$	206.7	202.0	1.01	−316.9	9.1	23.0	0.28	−3.3
Minimum variance	2.4	3.4	−0.07	7.6	3.4	5.3	0.13	5.5
Tangency	−49.3	51.4	−1.01	83.9	6.1	11.5	0.30	−0.2
Panel C: CAPM means								
$\gamma = -50$	2.7	0.3	0.13	2.6	2.7	0.3	0.13	2.6
$\gamma = -10$	2.8	1.2	0.13	2.4	2.8	1.2	0.13	2.4
$\gamma = -5$	2.9	2.1	0.13	2.3	2.9	2.1	0.13	2.2
$\gamma = -1$	3.5	6.4	0.13	1.4	3.5	6.4	0.13	1.4
$\gamma = 0$	4.3	12.8	0.13	0.2	4.0	10.5	0.13	0.2
$\gamma = 0.5$	5.9	25.6	0.13	−2.2	4.5	16.4	0.11	−0.8
Minimum variance	2.8	3.4	0.06	7.6	3.2	5.3	0.11	5.5
Tangency	3.6	7.7	0.13	1.2	3.6	7.8	0.13	1.1

Notes: A mean-variance efficient portfolio is defined in terms of an approximation to a power-utility function $u(w) = 1/\gamma w^{\gamma}$. The corresponding risk tolerance parameter employed in the mean-variance optimizer is $T = 1/(1-\gamma)$. The investment universe consists of twelve value-weighted industry groups. Quarterly portfolio revision with a 32-quarter estimation period is employed. In the absence of short-sales constraints, borrowing and lending are permitted at the riskfree lending rate. With short-sales constraints, the borrowing rate exceeds the lending rate, but the constrained tangency portfolio is constructed using the lending rate. Returns are measured in percent per quarter. The Sharpe ratio is defined as the mean excess return divided by the standard deviation of return. Figs. 1, 2, and 3 contain plots of the constrained (short sales precluded) frontiers and the unconstrained (short sales permitted) frontiers based on the dividend yield riskfree rate, historic, and CAPM means in panels A, B, and C, respectively.

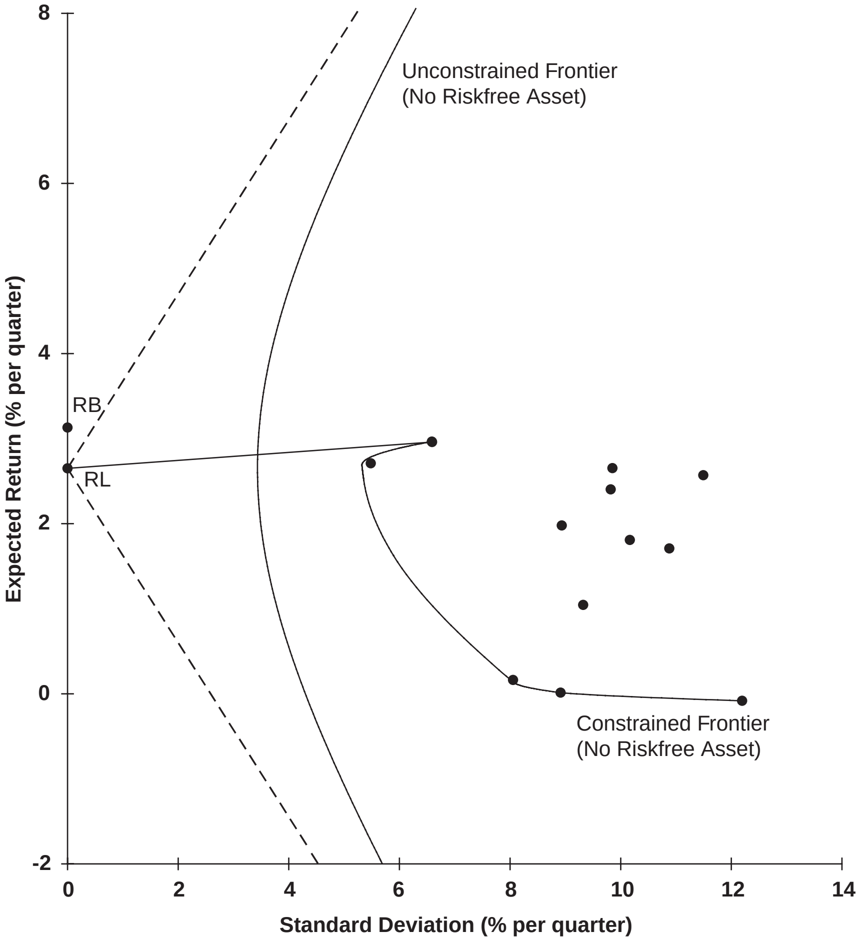


Fig. 1. Constrained and Unconstrained Minimum-Variance Frontiers Estimated from Dividend-Yield Riskfree-Rate Means in the Fourth Quarter of 1984. The Dots Represent the Expected Returns and Standard Deviations of the Twelve Industry Portfolios, Riskfree Lending (RL), and Riskfree Borrowing (RB). The Hyperbolic Curves Represent the Minimum-Variance Frontiers When There is no Riskfree Asset. The Dashed Lines Emanating from the RL Rate Represent the Unconstrained Minimum-Variance Frontier When RB and RL at the Same Rate are Added to the Opportunity Set. The Solid Line Emanating from the RL Rate Represents the Constrained Frontier When RB and RL at the Different Rates are Added to the Opportunity Set. In This Case, There is no Borrowing Rate Tangency. The Expected Return on the Tangency Portfolio on the Unconstrained Frontier Exceeds 12,500 Percent Per Quarter. Table 7 Contains the Expected Returns, Standard Deviations, Sharpe Ratios, and Realized Returns of the Minimum-Variance, Tangency, and Constant Risk Tolerance Portfolios.

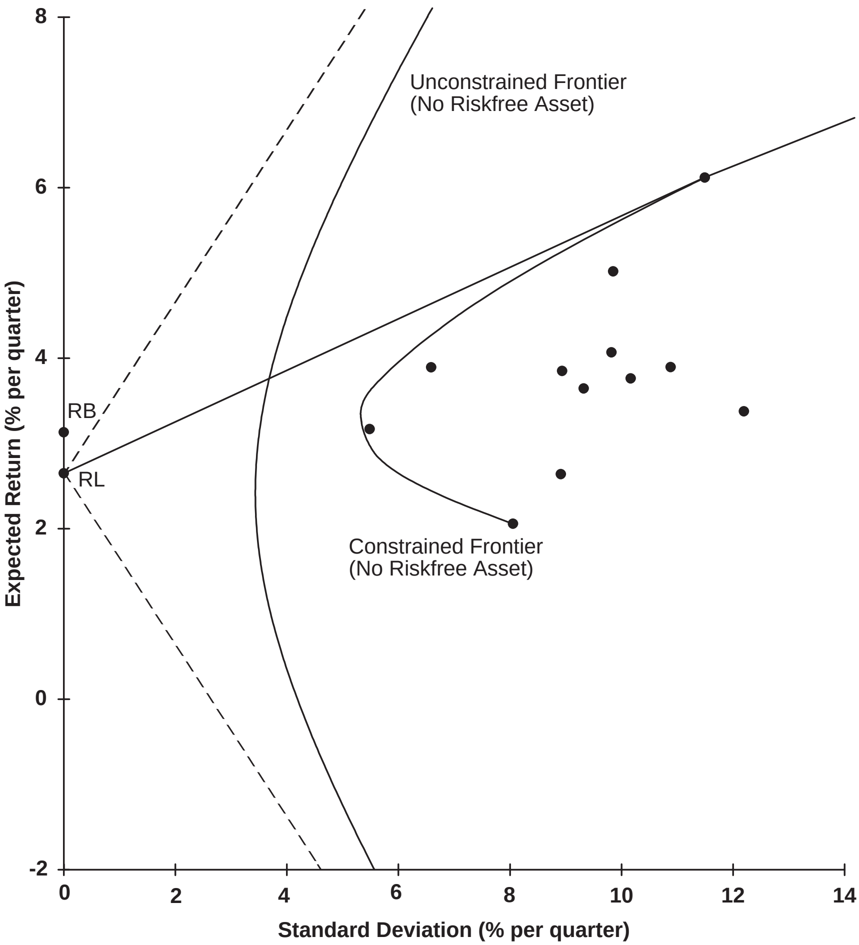


Fig. 2. Constrained and Unconstrained Minimum-Variance Frontiers Estimated from Historic Means in the Fourth Quarter of 1984. The Dots Represent the Expected Returns and Standard Deviations of the Twelve Industry Portfolios, Riskfree Lending (RL), and Riskfree Borrowing (RB). The Hyperbolic Curves Represent the Minimum-Variance Frontiers When There is no Riskfree Asset. The Dashed Lines Emanating from the RL Rate Represent the Unconstrained Minimum-Variance Frontier When RB and RL at the Same Rate are Added to the Opportunity Set. The Solid Lines Emanating from the RL and RB Rates Represent the Constrained Frontier When RB and RL at the Different Rates are Added to the Opportunity Set. The Expected Return of the Tangency Portfolio on the Unconstrained Frontier is Approximately –49 Percent Per Quarter. Table 7 contains the Expected Returns, Standard Deviations, Sharpe Ratios, and Realized Returns of the Minimum-Variance, Tangency, and Constant Risk Tolerance Portfolios.

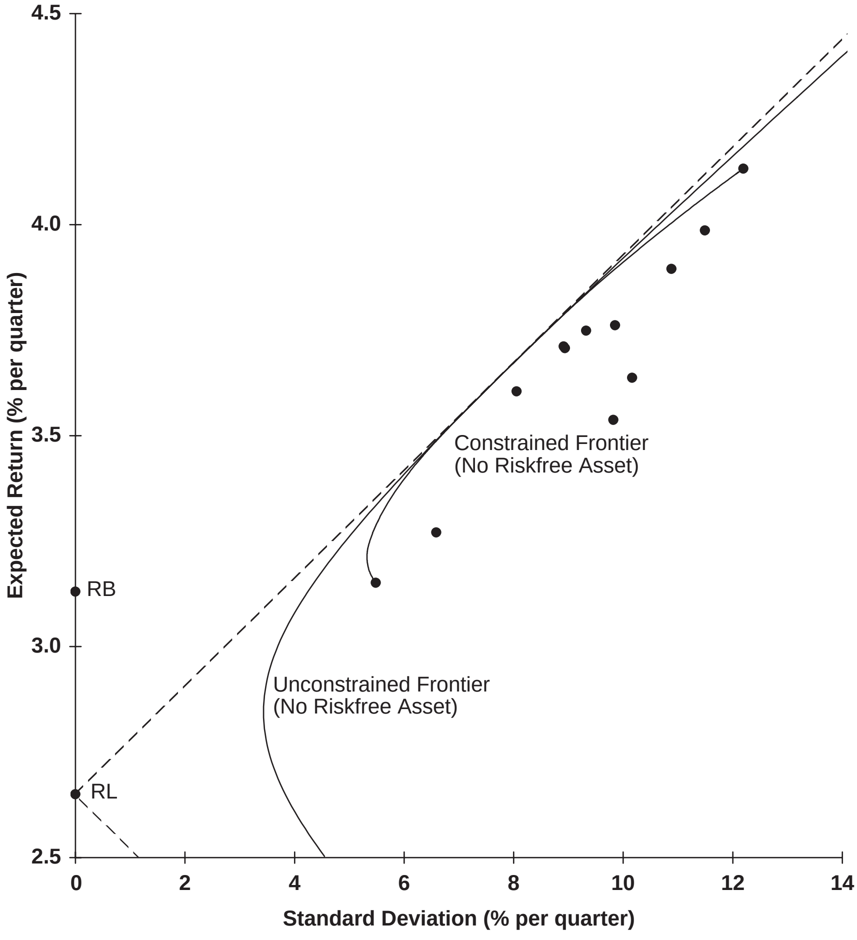


Fig. 3. Constrained and Unconstrained Minimum-Variance Frontiers Estimated from CAPM Means in the Fourth Quarter of 1984. The Dots Represent the Expected Returns and Standard Deviations of the Twelve Industry Portfolios, Riskfree Lending (RL), and Riskfree Borrowing (RB). The Hyperbolic Curves Represent the Minimum-Variance Frontiers When There is no Riskfree Asset. The Dashed Lines Emanating from the RL Rate Represent the Unconstrained Minimum-Variance Frontier When RB and RL at the Same Rate are Added to the Opportunity Set. Given that the Frontiers are so Close Together, the Constrained Frontier with Different RB and RL Rates is not Plotted. Table 7 Contains the Expected Returns, Standard Deviations, Sharpe Ratios, and Realized Returns of the Minimum-Variance, Tangency, and Constant Risk Tolerance Portfolios.

$T = T_{\mathrm{Tan}} = 1/(a - rc)$. Those investors whose risk tolerance parameters are less (greater) than $1/(a - rc)$ lend (borrow). In this case, the riskfree rate is infinitesimally less than the mean of the global minimum-variance portfolio, so the implied risk tolerance parameter of the investor who holds the unlevered tangency portfolio is approximately equal to 120. Consequently, all six MV-efficient portfolios lend. The lending ranges from 99.98 percent of wealth for the MV approximation to the -50 power, whose risk tolerance is approximately equal to 0.02, 98.33 percent of wealth for the MV approximation to the 0.5 power, whose risk tolerance is equal to 2. This reduces the expected return from over 12,500 percent to 4.7 and 213 percent, respectively.[14]

The frontiers and the asset means and standard deviations are shown in Fig. 1. With short sales permitted, it is somewhat hard to visualize that there is a tangency (with a mean of 12,500 percent) on the efficient frontier of risky assets. With short sales precluded, the minimum-variance frontier of risky assets takes on an unusual shape. The tangency portfolio (taken with respect to the lending rate) consists of a 100 percent investment in food and tobacco, and the efficient frontier does not extend past the tangency point, for example, no one borrows.

The characteristics of the historic mean portfolios in panel B and frontiers in Fig. 2 provide an interesting contrast. While the historic industry means do not differ dramatically from the DR means, the frontiers generated from them most certainly do. When short sales are permitted, the tangency portfolio is *inefficient* – its expected rate of return is -49 percent and its weights range from -9 to 16. On the contrary, the six MV-efficient portfolios plot on the upward sloping line (the efficient frontier) that emanates from the riskfree-lending rate. As noted in Note 7, MV investors with positive risk tolerance parameters achieve this position by shorting the tangency portfolio and investing over 100 percent of their wealth in riskfree lending. It is ironic, however, that the inefficient tangency portfolio realizes a large positive return, while two of the MV-efficient portfolios bankrupt. With short sales precluded, the frontiers and the ex ante and realized returns are much more realistic. Once again, the tangency portfolio consists of a 100 percent investment in one asset – this time the leisure industry.[15]

The CAPM mean results are presented in Table 7, panel C and in Fig. 3. Not surprisingly, the frontiers with and without short-sales constraints are almost indistinguishable. (The only reason the frontiers are distinguishable at all is that the scale of the vertical axis in Fig. 3 differs from the scale in Figs. 1 and 2.) In stark contrast to the investment policies based on the previous two sets of means, the more risk-tolerant MV-efficient investors lend, and the less risk-tolerant borrow, whether or not short-sales constraints are in effect. Moreover, the tangency portfolio contains eleven of the twelve industries when short-sales constraints are in effect.

7. SUMMARY AND CONCLUSIONS

This chapter compares the policies and performance of the global minimum-variance portfolio, the tangency portfolios, and six MV-efficient portfolios in an industry rotation setting, with and without short-sales constraints, when the means are estimated in four different ways. Although a number of researchers eschew estimating the means in light of the tangency portfolios' extreme characteristics, the results show that with minor exceptions, the passive policy of buying and holding the market portfolio outperforms the global minimum-variance portfolio, whether it is estimated with or without short-sales constraints. When short sales are permitted and means and investor risk tolerances are added to the analysis, the results are nothing short of bizarre. In the extreme, the tangency portfolio's weights are plus and minus thousands of times wealth! Ex ante many of the tangency portfolios are *inefficient*! Ex post the tangency portfolios – except those generated from the CAPM means – and most of the more risk-tolerant MV-efficient portfolios bankrupt! Although the results are easily explained in terms of the efficient set mathematics, to the best of my knowledge, they have not been documented in the asset allocation literature. Rather, they are obscured by typically reported performance measures. Arguably, portfolios generated from CAPM means perform the best. And, those generated from DR estimator means the worst. Yet, when short-sales constraints are imposed, just the opposite is true.

Although estimation error undoubtedly plays a part in determining the results, modeling error plays a more fundamental role. Active managers who hope to generate abnormal returns must employ means that have forecasting ability. But, given the acute sensitivity of the portfolio weights to small perturbations in equilibrium means, they must also impose short-sales constraints to avoid extreme positions and extreme outcomes. The empirical evidence shows that investors who utilize MV analysis without imposing short-sales constraints, without employing estimates of the means based on predictive variables, and without specifying their risk tolerance miss out on remarkably remunerative investment opportunities.

NOTES

1. Board and Sutcliffe (1994), Eichhorn, Gupta, and Stubbs (1998), and Clarke, de Silva, and Thorley (2002), among others, also examine the effects of imposing short-sales constraints on MV investment problems.

2. Surprisingly, in a number of instances, Chan et al. (1999) and Jaganathan and Ma (2003) estimate the sample covariance matrix with fewer observations than assets, which guarantees the covariance matrix will not be of full rank. If the covariance matrix is not of full rank and positive definite, then there is no guarantee that the solutions will be unique, and it may be possible to find a portfolio of risky assets that has a variance of zero, which in turn could lead to arbitrage opportunities. However in this case, any arbitrage opportunities would be illusionary because of the way the covariance matrix was estimated.

3. Solnik (1993), Klemkosky and Bharati (1995), Connor (1997), Beller, Kling, and Levinson (1998), Ferson and Siegel (2001), Marquering and Verbeek (2001), Fletcher and Hillier (2002), Avramov (2002, 2004), Avramov and Chordia (2006), and Avramov and Wermers (2006), among others, combine forecasts from information variables with MV optimization.

4. Black and Litterman (1992) employ the same argument to generate what they call equilibrium means.

5. The importance of short-sales constraints is implicit in these two basic characteristics of MV problems. By way of contrast, Jaganathan and Ma (2003) (Best & Grauer, 1990; Grauer, 1991) employ the Lagrange multipliers from constrained optimization to show how short-sales constraints are equivalent to shrinking the covariance matrix [the mean vector]. If we solve for the weights in the global minimum-variance portfolio with nonnegativity constraints imposed, the securities for which the nonnegativity constraint is binding will have positive Lagrange multipliers (dual variables) associated with them. Jaganathan and Ma show that if we shrink the covariance matrix using the multipliers and solve the minimization problem with the new covariance matrix and the budget constraint only, then we get the same answer. Specifically, whenever the nonnegativity constraint is binding for stock j, its covariances between stocks j and k are reduced by $\lambda_j + \lambda_k$ and its variance is reduced by $2\lambda_j$, where λ_j is the multiplier associated with asset j. Suppose instead that we solve the MV problem given by Eqs. (3)–(5), but without a riskless asset. Best and Grauer (1990) and Grauer (1991) show that if we add the multiplier divided by the risk tolerance parameter to the means of each stock that is shorted and solve the problem with the new means and the budget constraint only, then we get the same answer. This result is more general than Jaganathan and Ma's because it applies to any minimum-variance portfolio other than the global minimum-variance portfolio. But, given that there are different sets of binding short-sales constraints (and different sets of Lagrange multipliers) for each minimum-variance portfolio, I believe that it makes more sense to think of the short-sales constraints in terms of the extreme sensitivity of the solutions to any perturbation in the $(\mathbf{\Sigma}, \mathbf{x}_m)$ compatible means.

6. If the budget constraint is the only constraint, portfolio selection models permit investors to use the proceeds of unlimited short sales to finance unlimited long positions. However, the Federal Reserve Board's Regulation T does not allow investors to use the proceeds of short sales to finance long positions. Rather,

investors must leave the proceeds of short sales with the broker *and* must post margin to protect the broker in case the price goes up. Jacobs and Levy (2006) describe how large institutional investors can employ prime brokerage accounts that allow use of the proceeds of short sales. But the amounts one might short sell are not even of the same order of magnitude as some of the solutions documented in this chapter. Lintner (1965), Pastor (2000), Pastor and Stambaugh (2000), Avramov (2002) and Jacobs, Levy, and Markowitz (2005) consider MV problems where short positions are modeled in more realistic ways.

7. For illustrative purposes, I employ dividend yields and the riskfree rate as information variables in a predictive regression. I make no claim that this is in any sense optimal. Other information variables may lead to better results. Moreover, Pastor and Stambaugh (2006) argue that compared to predictive regressions, their new predictive systems deliver different and substantially more precise estimates of expected returns as well as different assessments of a given predictor's usefulness.

8. The efficient set mathematics can be found in Merton (1972), Roll (1977), and Huang and Litzenberger (1988). Best and Grauer (1990) state the efficient set mathematics in terms of the risk tolerance parameter.

9. Or stated in terms of risk aversion, the MV problem is $\max(\mu_p - 1/2A\sigma_p^2)$, where $A = 1/T$.

10. There is no practical way to take maintenance margins into account in my programs. In any case, it is evident from the results that they would come into play only for the more risk-tolerant strategies and even for them, only occasionally, and that the net effect would be relatively neutral.

11. In a power-utility setting, Grauer (2008a) employs DR, DRH, and DRCAPM means, which are DR means "shrunk" to CAPM means as suggested by Black and Litterman (1992). The portfolios generated from the DRCAPM means outperform the portfolios generated from the DRH means, which in turn outperform the portfolios generated from the DR means.

12. As noted, if the riskfree return is greater than the expected return of the global minimum-variance portfolio, the tangency portfolio is on the inefficient portion of the minimum-variance frontier of risky assets. The *inefficient* frontier (characterized by negative values of T and negative Sharpe ratios) is a downward sloping line from the riskfree asset through the tangency portfolio. The *efficient* frontier (characterized by positive values of T and positive Sharpe ratios) is an upward sloping line emanating from the riskfree asset. The efficient frontier is traced out by shorting the tangency portfolio and investing the proceeds in the riskfree asset. See the frontier in Fig. 2.

13. Michaud (1989) notes that the use of sample estimators in the optimization process has a tendency to maximize the effects of errors in the inputs. That appears to be the case in Table 2 where, for the most part, the average values of the portfolios' ex ante means and Sharpe ratios exceed their (single-valued) ex post counterparts, and the average values of the portfolios' ex ante standard deviations are less than the ex post standard deviations. But, an earlier version of the paper, which analyzed eight sets of means, showed that different sets of parameters in Best and Grauer's (1985) $(\boldsymbol{\Sigma}, \mathbf{x}_m)$ compatible means can eliminate, exaggerate, or completely reverse the differences in these ex post and ex ante return variables – all the while yielding exactly the same tangency portfolio.

14. One of the more appealing aspects of power utility with <1 is that with discrete joint return distributions, there is no ex ante probability of bankruptcy. The MV approximation to power utility works well with quarterly decision horizons when short sales are precluded. But it breaks down here when short sales are permitted. In results not reported in the text, I confirmed that with either DR or historic means and 32 states of nature corresponding to the moving 32-quarter window employed as the joint return distribution as in the Grauer–Hakansson power-utility papers, the logarithmic utility portfolio does not bankrupt in any of the ex ante states of nature. But, the MV-approximation to the logarithmic policy bankrupts in many of these same states.

15. Interestingly, with riskfree lending and borrowing at different rates, the efficient frontier contains a "kink" at the tangency point. Moreover, the investors' holdings of the riskfree assets are completely different than when short sales are permitted. Although it is not recorded in the table, the MV portfolios that approximate the –50 to –5 powers lend, those that approximate powers –1 and greater borrow, and a –3 power investor neither borrows nor lends.

ACKNOWLEDGMENTS

I thank the Social Sciences Research Council of Canada for financial support, Peter Klein, Geoff Poitras, and Daniel Smith for many helpful discussions, and Christopher Fong for most capable assistance.

REFERENCES

Avramov, D. (2002). Stock return predictability and model uncertainty. *Journal of Financial Economics*, *64*, 423–458.

Avramov, D. (2004). Stock return predictability and asset pricing models. *Review of Financial Studies*, *17*, 699–738.

Avramov, D., & Chordia, T. (2006). Predicting stock returns. *Journal of Financial Economics*, *82*, 387–415.

Avramov, D., & Wermers, R. (2006). Investing in mutual funds when returns are predictable. *Journal of Financial Economics*, *81*, 339–377.

Beller, K. R., Kling, J. L., & Levinson, M. (1998). Are industry stock returns predictable? *Financial Analysts Journal*, *54*(September/October), 42–57.

Best, M. J., & Grauer, R. R. (1985). Capital asset pricing compatible with observed market value weights. *Journal of Finance*, *40*(March), 85–103.

Best, M. J., & Grauer, R. R. (1990). The efficient set mathematics when mean-variance problems are subject to general linear constraints. *Journal of Economics and Business*, *42*(May), 105–120.

Best, M. J., & Grauer, R. R. (1991). On the sensitivity of mean-variance-efficient portfolios to changes in asset means: Some analytical and computational results. *Review of Financial Studies*, *4*, 315–342.

Best, M. J., & Grauer, R. R. (1992). Positively weighted minimum-variance portfolios and the structure of asset expected returns. *Journal of Financial and Quantitative Analysis*, *27*(December), 513–537.

Black, F., & Litterman, R. (1992). Global portfolio optimization. *Financial Analysts Journal*, *48*(October), 28–43.

Board, J. L. G., & Sutcliffe, C. M. S. (1994). Estimation methods in portfolio selection and the effectiveness of short sales restrictions: U.K. evidence. *Management Science*, *40*(April), 516–534.

Chan, L. K., Karceski, J., & Lakonishok, J. (1999). On portfolio optimization: Forecasting covariances and choosing the risk model. *Review of Financial Studies*, *12*(Winter), 937–974.

Clarke, R., de Silva, H., & Thorley, S. (2002). Portfolio constraints and the fundamental law of active management. *Financial Analysts Journal*, *58*(September/October), 48–66.

Connor, G. (1997). Sensible return forecasting for portfolio management. *Financial Analysts Journal*, *53*(September/October), 44–51.

Efron, B., & Morris, C. (1973). Stein's estimation rule and its competitors – an empirical Bayes approach. *Journal of the American Statistical Association*, *68*, 117–130.

Efron, B., & Morris, C. (1975). Data analysis using Stein's estimator and its generalizations. *Journal of the American Statistical Association*, *70*, 311–319.

Efron, B., & Morris, C. (1977). Stein's paradox in statistics. *Scientific American*, *236*, 119–127.

Eichhorn, D., Gupta, F., & Stubbs, E. (1998). Using constraints to improve the robustness of asset allocation. *Journal of Portfolio Management*, *24*(Spring), 41–48.

Eun, C. S., & Resnick, B. G. (1994). International diversification of investment portfolios: U.S. and Japanese perspectives. *Management Science*, *40*(January), 140–161.

Ferson, W. E., & Siegel, A. F. (2001). The efficient use of conditioning information in portfolios. *Journal of Finance*, *56*(June), 967–982.

Fletcher, J., & Hillier, J. (2002). On the usefulness of linear factor models in predicting expected returns in mean-variance analysis. *International Review of Financial Analysis*, *4*, 449–466.

Frost, P. A., & Savarino, J. E. (1988). For better performance: Constrain portfolio weights. *Journal of Portfolio Management*, *14*, 29–34.

Grauer, R. R. (1991). Further ambiguity when performance is measured by the security market line. *Financial Review*, *26*(November), 569–585.

Grauer, R. R. (2008a). On the predictability of stock market returns: Evidence from industry-rotation strategies. *Journal of Business and Management*, *14*, 47–71.

Grauer, R. R. (2008b). Benchmarking measures of investment performance with perfect-foresight and bankrupt asset allocation strategies. *Journal of Portfolio Management*, *34*(Summer), 43–57.

Grauer, R. R., & Hakansson, N. H. (1986). A half-century of returns on levered and unlevered portfolios of stocks, bonds, and bills, with and without small stocks. *Journal of Business*, *59*(April), 287–318.

Grauer, R. R., & Hakansson, N. H. (1987). Gains from international diversification: 1968–85 returns on portfolios of stocks and bonds. *Journal of Finance*, *42*(July), 721–739.

Grauer, R. R., & Hakansson, N. H. (1993). On the use of mean-variance and quadratic approximations in implementing dynamic investment strategies: A comparison of the returns and investment policies. *Management Science*, *39*(July), 856–871.

Grauer, R. R., & Hakansson, N. H. (1995). Stein and CAPM estimators of the means in asset allocation. *International Review of Financial Analysis*, *4*, 35–66.

Grauer, R. R., Hakansson, N. H., & Shen, F. C. (1990). Industry rotation in the U.S. stock market: 1934–1986 returns on passive, semi-passive, and active strategies. *Journal of Banking and Finance*, *14*(August), 513–535.

Huang, C., & Litzenberger, R. H. (1988). *Foundations for financial economics*. North-Holland, New York.

Jacobs, B. I., & Levy, K. N. (2006). Enhanced active equity strategies: Relaxing the long-only constraint in the pursuit of active return. *Journal of Portfolio Management*, *32*(Spring), 45–55.

Jacobs, B. I., Levy, K. N., & Markowitz, H. M. (2005). Portfolio optimization with factors, scenarios, and realistic short positions. *Operations Research*, *53*(July–August), 586–599.

Jaganathan, R., & Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance*, *58*(August), 1651–1683.

Jobson, J. D., & Korkie, B. (1980). Estimation for Markowitz efficient portfolios. *Journal of the American Statistical Association*, *75*(September), 544–554.

Jobson, J. D., & Korkie, B. (1981). Putting Markowitz theory to work. *Journal of Portfolio Management*, *7*(Summer), 70–74.

Jobson, J. D., Korkie, B., & Ratti, V. (1979). Improved estimation for Markowitz portfolios using James-Stein type estimators. *Proceedings of the American Statistical Association, Business and Economics Statistics Section*, *41*, 279–284.

Jorion, P. (1985). International portfolio diversification with estimation risk. *Journal of Business*, *58*(July), 259–278.

Jorion, P. (1986). Bayes-Stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis*, *21*(September), 279–292.

Jorion, P. (1991). Bayesian and CAPM estimators of the means: Implications for portfolio selection. *Journal of Banking and Finance*, *15*, 717–727.

Klemkosky, R. C., & Bharati, R. (1995). Time-varying expected returns and asset allocation. *Journal of Portfolio Management*, *21*(Summer), 80–88.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics*, *47*(December), 13–47.

Markowitz, H. (1959). *Portfolio selection: Efficient diversification of investments*. New York: Wiley.

Marquering, W., & Verbeek, M. (2001). The economic value of predicting stock index returns and volatility. *Journal of Financial and Quantitative Analysis*, *39*, 407–429.

Merton, R. C. (1972). An analytic derivation of the efficient frontier. *Journal of Financial and Quantitative Analysis*, *10*(September), 1851–1872.

Merton, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica*, *41*(September), 867–888.

Michaud, R. O. (1989). The Markowitz optimization enigma: Is 'optimized' optimal? *Financial Analysts Journal*, *45*(January/February), 31–42.

Pastor, L. (2000). Portfolio selection and asset pricing models. *Journal of Finance*, *55*(February), 179–223.

Pastor, L., & Stambaugh, R. (2000). Comparing asset pricing models: An investment perspective. *Journal of Financial Economics*, *56*, 323–361.

Pastor, L., & Stambaugh, R. (2006). *Predictive systems: Living with imperfect predictors*. CRSP Working Paper no. 617.

Roll, R. (1977). A critique of the asset pricing theory's tests; part I: On past and potential testability of the theory. *Journal of Financial Economics*, *4*, 129–176.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, *19*(September), 425–442.

Sharpe, W. F. (1970). *Portfolio theory and capital markets*. New York: McGraw-Hill.

Solnik, B. (1993). The performance of international asset allocation strategies using conditioning information. *Journal of Empirical Finance*, *1*(June), 33–55.

Stein, C. (1955). *Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. Proceedings of the 3rd Berkeley Symposium on Probability and Statistics I.* Berkeley: University of California Press.

A FUZZY PROGRAMMING APPROACH TO FINANCIAL PORTFOLIO MODEL

Kenneth D. Lawrence, Dinesh R. Pai,
Ronald K. Klimberg and Sheila M. Lawrence

ABSTRACT

The Black and Litterman model (1992) for estimating asset returns is widely used in industry and has been widely studied in the academic and professional literature. Black and Litterman offer a way to incorporate investor's views into asset-pricing. This chapter provides a description of the Black and Litterman model. The model is analyzed using fuzzy goal programming approach using appropriate membership functions. We consider a real world financial example to implement our approach.

INTRODUCTION

The Black and Litterman model (1992) was first published by Fischer Black and Robert Litterman of Goldman Sachs in an internal Goldman Sachs Fixed Income document in 1990. The model makes two significant contributions to the problem of asset allocation. First, it provides an intuitive prior, the CAPM equilibrium market portfolio, as a starting point

Financial Modeling Applications and Data Envelopment Applications
Applications of Management Science, Volume 13, 53–59

ISSN: 0276-8976/doi:10.1108/S0276-8976(2009)0000013005

for the estimation of asset returns. Second, the Black and Litterman model provides a clear way to specify investors' views and to blend the investors' views with prior information. These views can be partial or complete, and the views can span arbitrary and overlapping sets of assets. The model estimates expected excess returns and covariances, which can be used as input to an optimizer. Combining these two contributions results in a new vector of expected returns. This improved vector of expected returns can then be used in the portfolio optimization process. In this chapter, we compare the results of the Black and Litterman model using fuzzy goal programming approach with the appropriate membership function (Sharpe, 1971; Zimmermann, 1991).

BLACK AND LITTERMAN MODEL

The Black–Litterman model creates stable, mean-variance efficient portfolios, based on an investor's unique insights, which overcome the problem of input-sensitivity. According to Lee (2000), the Black–Litterman model, also, "largely mitigates" the problem of estimation error-maximization (Michaud, 1989) by spreading the errors throughout the vector of expected returns.

The Black and Litterman model combines views of the investor and the market equilibrium on the expected return of assets in one formula. This formula should be a better approximation of the expected returns. These expected returns, or more precisely the estimator of the expected return, could then be used in a mean-variance optimizer (Karacabey, 2006).

The resulting distribution for $E(r)$ is a multiplicative normal distribution with mean:

$$\left[(\tau\Sigma)^{-1} + P'\Omega^{-1}P\right]^{-1}\left[(\tau\Sigma)^{-1}\pi + P'\Omega^{-1}q\right] \tag{1}$$

Black and Litterman model use $E(r)$ as the vector of expected returns in the Markowitz model. The Black and Litterman model can be summarized by the following points (Karacabey, 2006):

1. The market consists of n assets. The assets have a return $r \in \mathbb{R}^n$, with variance Σ and expected return $E(r)$. The expected return $E(r)$ is an unknown and normally distributed random variable; it is assumed to have mean π and variance $\tau\Sigma$.
2. The first source of information about $E(r)$ is the equilibrium returns π. The equilibrium returns are found by $\pi = \delta\Sigma w_m$, where δ is a (world) risk aversion coefficient or a ratio of the (world) market portfolio and w_m are

the market weights. They can be represented as $\pi = E(r) + u$, with $u \sim N(0, \tau\Sigma)$, where τ is some proportionality constant.
3. The second source of information is the k views of the investor. The views are expressed as $PE(r) = q + \varepsilon$, where $P \in \mathbb{R}^{nxn}$, $q \in \mathbb{R}^n$, and $\varepsilon \sim N(0, \Omega)$, Ω is a diagonal $(k \times k)$ matrix.
4. Combination of these two sources of information leads to $E(r)$ being normally distributed with mean $\left[(\tau\Sigma)^{-1} + P'\Omega^{-1}P\right]^{-1}\left[(\tau\Sigma)^{-1}\pi + P'\Omega^{-1}q\right]$ and variance $\left[(\tau\Sigma)^{-1} + P'\Omega^{-1}P\right]^{-1}$.
5. This mean can be used in a mean-variance optimization process to obtain a mean-variance efficient portfolio.

Example. We take the following example to illustrate the Black and Litterman model.

The expected returns on stocks, bonds, and money market are

	Stocks	Bonds	Money Market
Market rate of return	10.73%	7.37%	6.27%

By applying formula in Eq. (1) to compute expected returns $E(r)$, we get

	Stocks	Bonds	Money Market
Market rate of return	12.00%	7.50%	2.50%

The covariance matrix is given as

Covariance Matrix

	Stocks	Bonds	Money market
Stocks	0.0278		
Bonds	0.0039	0.0111	
Money market	0.0002	−0.0002	0.0012

We solve the non-linear programming problem as follows (Stone, 1973):

Minimize

$$0.002778x_S^2 + (2)(0.00387)x_S x_B + (2)(0.00021)x_S x_M + 0.01112x_B^2 - (2)(0.00020)x_B x_M + 0.00115x_M^2$$

$$\text{Subject to :} \quad \begin{aligned} & 0.12x_S + 0.075x_B + 0.0225x_M \geq R \\ & x_S + x_B + x_M = 1 \\ & x_S \geq 0, \quad x_B \geq 0, \quad x_M \geq 0 \end{aligned}$$

Where, x_S, x_B, x_M are the proportion of investments in stocks, bonds, and money market.

Solving for $R = 6.5\%$, we get the following solution: Variance $= 0.0038$, $x_S = 0.251, \quad x_B = 0.324, \quad x_M = 0.425$

FUZZY GOAL PROGRAMMING (WITH TRIANGULAR MEMBERSHIP FUNCTION)

For the fuzzy goal programming approach, we use the beta (β) of portfolio as a measure of risk. We calculate β for each instrument using, for example, the following formula:

$$\beta_{\text{Stocks}} = \frac{\text{Cov}\left(r_{\text{Stocks}}, r_{\text{MoneyMarket}}\right)}{\text{Var}\left(r_{\text{MoneyMarket}}\right)}$$

Let us assume that both the primary (risk) and the secondary (returns) goals are fuzzy. For the risk goal, we assume that the portfolio has a risk index β of around 0.05, with a tolerance limit of [0.05, 0.05]. For the returns goal, we assume that the portfolio will provide an annual return of around 6.5%, with a tolerance limit of [1.0%, 1.0%].

Let $G_k(x)$ denote the kth fuzzy goal with a triangular membership function

$$\mu_i(x) = \frac{G_i(x) - L_i}{g_i - L_i} \qquad \text{for some } i$$

$$\mu_i = \frac{U_i - G_i(x)}{U_i - g_i} \qquad \text{for some } i$$

β_s	β_b	β_m
0.1796	0.1392	−0.1693

Then the resulting linear programming formulation is

$$
\begin{aligned}
\text{Maximize :}\quad & Z = \lambda \\
\text{Subject to :}\quad & \lambda \le \frac{(0.1796x_{\mathrm{S}} + 0.1392x_{\mathrm{B}} - 0.1693x_{\mathrm{M}}) - 0.00}{0.05 - 0.00} \quad [\text{risk } (\beta) \text{ constraint}] \\
& \le 3.6x_{\mathrm{S}} + 2.8x_{\mathrm{B}} - 3.4x_{\mathrm{M}}, \\
& \lambda \le \frac{0.1 - (0.1796x_{\mathrm{S}} + 0.1392x_{\mathrm{B}} - 0.1693x_{\mathrm{M}})}{0.1 - 0.05} \quad [\text{risk } (\beta) \text{ constraint}] \\
& \le 2 - 3.6x_{\mathrm{S}} - 2.8x_{\mathrm{B}} + 3.4x_{\mathrm{M}}, \\
& \lambda \le \frac{(0.12x_{\mathrm{S}} + 0.075x_{\mathrm{B}} + 0.025x_{\mathrm{M}}) - 0.055}{0.065 - 0.055} \quad (\text{risk constraint}) \\
& \le 12x_{\mathrm{S}} + 7.5x_{\mathrm{B}} + 2.5x_{\mathrm{M}} - 5.5, \\
& \lambda \le \frac{0.075 - (0.12x_{\mathrm{S}} + 0.075x_{\mathrm{B}} + 0.025x_{\mathrm{M}})}{0.075 - 0.065} \quad (\text{risk constraint}) \\
& \le 7.5 - 12x_{\mathrm{S}} - 7.5x_{\mathrm{B}} - 2.5x_{\mathrm{M}}, \\
& x_{\mathrm{S}} + x_{\mathrm{B}} + x_{\mathrm{M}} = 1 \quad (\text{total investment constraint}) \\
& x_j, \lambda \ge 0, \quad j = 1, 2.
\end{aligned}
$$

The solver finds following optimal solution:

$$x_{\mathrm{S}}^* = 0.117, \quad x_{\mathrm{B}}^* = 0.577, \quad x_{\mathrm{M}}^* = 0.305, \quad \lambda = 1.0$$

FUZZY GOAL PROGRAMMING (WITH TRAPEZOIDAL MEMBERSHIP FUNCTION)

For the fuzzy goal programming approach (with Trapezoidal Membership Function), we use the beta (β) of portfolio as a measure of risk. We calculate β for each instrument using, for example, the following formula:

$$\beta_{\mathrm{Stocks}} = \frac{\mathrm{Cov}\left(r_{\mathrm{Stocks}}, r_{\mathrm{MoneyMarket}}\right)}{\mathrm{Var}\left(r_{\mathrm{MoneyMarket}}\right)}$$

Let us assume that both the primary (risk) and the secondary (returns) goals are fuzzy. Suppose we are not sure about the best value or point estimate of the risk index and annual returns. In this case we assume an interval estimate of 0.025–0.05 for the risk index and an interval estimate of 6.0–6.5 for the annual returns.

In this case we use a trapezoidal membership function to formulate this fuzzy goal programming model.

Let $G_k(x)$ denote the kth fuzzy goal with a trapezoidal membership function

$$\begin{aligned}
\mu_i(x) &= \frac{G_i(x) - b_1}{b_2 - b_1}, && \text{for } b_1 \le G_i(x) \le b_2 \\
\mu_i(x) &= \frac{b_4 - G_i(x)}{b_4 - b_3}, && \text{for } b_3 \le G_i(x) \le b_4 \\
\mu_i(x) &= 0, && \text{otherwise}
\end{aligned}$$

Then the resulting linear programming formulation is

$$\begin{aligned}
\text{Maximize}: \quad & Z = \lambda \\
\text{Subject to}: \quad & \{\text{for risk index constraint:} \\
& b_1 = 0.0, \quad b_2 = 0.025, \\
& b_3 = 0.05, \quad b_4 = 0.1\} \\
& \lambda \le \frac{(0.1796x_S + 0.1392x_B - 0.1693x_M) - 0.00}{0.025 - 0.00} \quad [\text{risk } (\beta) \text{ constraint}] \\
& \le 7.18x_S + 5.57x_B - 6.77x_M, \\
& \lambda \le \frac{0.1 - (0.1796x_S + 0.1392x_B - 0.1693x_M)}{0.1 - 0.05} \quad [\text{risk } (\beta) \text{ constraint}] \\
& \le 2 - 3.59x_S - 2.78x_B + 3.39x_M, \\
& \{\text{for returns constraint} \\
& b_1 = 5.5, \quad b_2 = 6.00, \\
& b_3 = 6.5, \quad b_4 = 7.5\} \\
& \lambda \le \frac{(0.12x_S + 0.075x_B + 0.025x_M) - 0.055}{0.06 - 0.055} \quad (\text{returns constraint}) \\
& \le 24x_S + 15x_B + 5x_M - 11.0, \\
& \lambda \le \frac{0.075 - (0.12x_S + 0.075x_B + 0.025x_M)}{0.075 - 0.065} \quad (\text{returns constraint}) \\
& \le 7.5 - 12x_S - 7.5x_B - 2.5x_M, \\
& x_S + x_B + x_M = 1 \quad (\text{total investment constraint}) \\
& x_j, \quad \lambda \ge 0, \quad j = 1, 2
\end{aligned}$$

The solver finds following optimal solution:

$$x_S^* = 0.1, \quad x_B^* = 0.544, \quad x_M^* = 0.356, \quad \lambda = 1.34$$

Table 1. Comparison of Three Methods.

Investments	Non-Linear	Fuzzy GP Triangular	Fuzzy GP Trapezoidal
Stocks (X_S)	0.251	0.117	0.100
Bonds (X_B)	0.324	0.577	0.544
Money Market (X_M)	0.425	0.305	0.356

RESULTS AND CONCLUSIONS

The results for the three methods discussed in this chapter are shown in Table 1. The table shows the proportion of investment in each of the instruments. Overall, all the three methods are comparable; however, the fuzzy methods give conservative results with a low proportion of investments in volatile instrument, that is, stocks and a higher proportion of investments in relatively stable instruments such as Bonds and Money market.

In this chapter, we compare the Black and Litterman optimization model using fuzzy expected returns. Using the fuzzy expected returns helps in minimizing the total variance compared with the original model. Moreover, using trapezoidal membership functions helps further in minimizing the total variance compared with the triangular membership functions.

REFERENCES

Black, F., & Litterman, R. (1992). Global portfolio optimization. *Financial Analysts Journal*, *48*(5), 28–43.

Karacabey, A. (2006). Is mean variance efficient than MAD in Istanbul? *International Research Journal of Finance and Economics*, *3*, 114–120.

Lee, W. (2000). *Advanced theory and methodology of tactical asset allocation*. New York: Wiley.

Michaud, R. (1989). The Markowitz optimization enigma: 'Is optimized' optimal? *Financial Analysts Journal*, *45*(1), 31–42.

Sharpe, W. (1971). A linear programming approximation for the general portfolio analysis problem. *Journal of Financial and Quantitative Analysis*, 263–1275.

Stone, B. (1973). A linear programming formulation of the general portfolio selection problem. *Journal of Financial and Quantitative Analysis*, *8*, 621–636.

Zimmermann, H. J. (1991). *Fuzzy set theory and its applications*. Boston, MA: Kluwer Academic Publishers.

BANKRUPTCY PREDICTION IN RETAIL INDUSTRY USING LOGISTIC REGRESSION

Kenneth D. Lawrence, Dinesh R. Pai and Gary Kleinman

ABSTRACT

In view of the failure of high-profile companies such as Circuit City and Linens n Things, Financial distress or bankruptcy prediction of retail and other firms has generated much interest recently. Recent economic conditions have led to predictions of a wave of retail bankruptcies (e.g., McCracken and O'Connell, 2009). This research develops and tests a model for the prediction of bankruptcy of retail firms. We use accounting variables such as inventories, liabilities, receivables, net income (loss), and revenue. Some guiding discriminate rule is given, and a few factors were identified as measures of a profitable company.

INTRODUCTION

Managers have been grappling with the problem of extracting patterns out of the vast databases generated by their systems. The advent of powerful information systems in organizations and the consequent agglomeration of

Financial Modeling Applications and Data Envelopment Applications
Applications of Management Science, Volume 13, 61–69

ISSN: 0276-8976/doi:10.1108/S0276-8976(2009)0000013006

vast pools of data since the mid-1980s have created renewed interest in the usefulness of discriminant analysis (DA). Expert systems have come to the aid of managers in their day-to-day decision making with many successful applications in financial planning, sales management, and other areas of business operations (Erenguc & Koehler, 1990).

Since Fisher's (1936) seminal work on linear DA, numerous methods have been developed for classification purposes. DA has been successfully applied in many business applications including building credit scoring models for predicting credit risk and investigating product failures (Dillon, Calantone, & Worthing, 1979; Myers & Forgy, 1963). Logistic regression is a related statistical method which is now widely used and Westin (1973) was one of the first to apply it in a binary choice situation.

In general, classification models assign observations of unknown class membership to a number of specified classes or groups using a set of explanatory variables associated with the group. These models have found myriad business applications such as in credit evaluation systems (Myers & Forgy, 1963), differentiating bank charge-card holders (Awh & Waters, 1974), screening credit applicants (Capon, 1982), assessing project implementation risk (Anderson & Narasimhan, 1979), predicting consumer innovators for new product diffusion (Robertson & Kennedy, 1968), predicting corporate bankruptcy (Altman, 1968), investigating new product success or failure (Dillon et al., 1979), predicting bank failures (Tam & Kiang, 1992), and approving loan applications (Gallant, 1988). These models have been particularly useful in market segmentation based on observable and product-specific bases. The advances in computers and information technology have further increased the efficacy of such approaches whereby vast amounts of historical customer data can be processed to understand customer needs and wants. This has resulted in more focused marketing strategies that result in lower costs, higher response rates, and consequently higher profits (Zahavi & Levin, 1997). The usefulness of bankruptcy prediction models to auditors has been shown in Sun (2007). Auditors have a legal responsibility, under Generally Accepted Accounting Standards, to evaluate the going concern status of their auditing clients. Sun examined the relative effectiveness of differently formulated bankruptcy prediction models presented in the literature and compared them to an audit firm–related model. Sun found that the audit-related model used in predicting going concern opinions was not as effective as a hazard model.

Ghargori, Chan, and Faff (2006) compare options-based to accounting-based approaches in the assessment of default risk. They found that options-based models, based on the work of Merton (1974: cited by Ghargori et al.), outperformed the accounting-based models. Ghargori et al. note that accounting-based default measures have two problems. These problems stem from the fact that financial statements are backward looking, that is, they describe what happened previously and assume the firm's continuance as a viable entity. Ghargori et al. also note that leverage ratios and other accounting measures differ between industries and that the relationship between various accounting numbers and default risk is, as they described it, "intuitive," not grounded on solid theory. The authors fail to note that accounting methods vary between industries, as well as within industries, introducing an additional source of variance into the data typically used. Regrettably, there is no obvious cure to the latter problem. Using the options-based model, whether path dependent or path independent, however, is problematic for practitioners in the business or allied fields since the construction of the models is very complex technically, relying as it does on understanding the distributional assumptions underlying the data used.[1]

In this chapter, we predict the bankruptcy status of retail companies based on their financial variables. Unlike option or more sophisticated models, the data used in this study are easily obtainable and intuitively understandable to the finance or accounting practitioner. Although process models of firm failure are also possible (e.g., Ooghe & De Prijcker, 2008), we find these to be interesting but impossible to use for firm samples of any size. By choosing only firms from a single industry, we eliminate problems caused using cross-sectional data (e.g., variations in systematic causes of bankruptcy risk versus idiosyncratic causes; see Parnes, 2009) at the expense of losing generalizability. We eschew such complex models as being of limited usefulness to the finance or accounting practitioner. For the purpose of this exercise, we select a list of 70 retail companies based in North America.

LOGISTIC REGRESSION

Logistic regression, a statistical modeling method for categorical data, has expanded from its origins in biomedical research to fields such as business and finance, engineering, marketing, economics, and health policy Meyers, Gamst, and Guarino (2005). The availability of sophisticated statistical software and high-speed computing has further increased the utility of logistic regression as an important statistical tool.

Logistic regression is particularly suitable for estimating categorical (dichotomous or polytochomous) dependent variables using maximum likelihood estimation (MLE) procedure. Logistic regression models use MLE as their convergence criterion. Logistic regression allows one to predict such dichotomous outcomes as presence/absence, success/failure, buy/don't buy, default/don't default, and survive/die. The independent variables may be categorical, continuous, or a combination of the both. We can think of categorical variables as dividing the observations into several classes. For example, if Y denotes a recommendation on holding/selling/buying a stock, then we have a categorical variable with three categories. We can think of each stock in the dataset as belonging to one of the three classes: the "hold" class, the "sell" class, and the "buy" class. Logistic regression has found two broad applications in applied research: classification (predicting group membership) and profiling (differentiating between two groups based on certain factors) (Tansey, White, & Long, 1996; Shmueli, Patel, & Bruce, 2006).

In general, the logistic regression model has the form

$$\log\left[\frac{p}{1-p}\right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n = x\beta \tag{1}$$

where p is the probability of the outcome of interest, β_0 an intercept term, β_i the coefficient associated with the corresponding dependent (explanatory) variable x_i, $x = (1, x_1, x_2, \ldots, x_n)$ and $\beta = (\beta_0, \beta_1, \ldots, \beta_n)'$.

The probability of the outcome of interest, p, is expressed as a non-linear function of the predictors in the form

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n)}} \tag{2}$$

Eq. (2) ensures that the right hand side will always lead to values within the interval [0, 1]. This is called the *logistic response function*.

In Eq. (1), the expression

$$\frac{p}{1-p} = \text{odds}, \quad \text{which can be rewritten as } p = \frac{\text{odds}}{1 + \text{odds}} \tag{3}$$

Hence, in logistic regression, one estimates the log of probability odds, also known as the *logit,* by a linear combination of the predictor variables. The *logit* takes on values from $-\infty$ to $+\infty$.

Taking exponentials of both sides of Eq. (1) leads to

$$p = \frac{e^{x\beta}}{1 + e^{x\beta}} \tag{4}$$

DATA AND RESULTS

Financial data of the publicly listed retail companies in North America were collected for the year 2006. The data was collected largely from COMPUSTAT. In addition, we collected the data about bankrupt firms from the Web BRD, UCLA data base. We started with a consideration set of 85 publicly listed North American retailers. However, due to lack of financial data for some years, or a merger, our final set reduced to 70 companies.[2] The descriptive statistics of the firms in our sample are given in Table 1.

Privately held retailers were not considered for this study due to unavailability of their financial data. The status of a retailer is represented by a binary variable: a bankrupt retailer is labeled as "1" and non-bankrupt retailer is labeled as "0." We used a combination of variables to arrive at our regression model and found that the model with the following variables, inventories, liabilities, receivables, net income (loss), and revenue, gave us good results. These results make intuitive sense. For example, high levels of inventories carry obvious risks of inventory obsolescence and therefore loss. Heavy liabilities relative to other firms suggest the burden of servicing the debt and raise issues as to how the liabilities will be satisfied (Tunick, 2002). Receivable balances suggest potential cash flow difficulties, given that customers may not be paying amounts due on time (Ketzner, 2005). Net income or loss as a variable provides a quick, easily understandable index of

Table 1. Descriptive Statistics of the Financial Variables Used in This Exercise.

Variables	Mean	SD	Kurtosis	Skewness	Range	Count
Inventories	2,123,162	3,014,975	3.80	2.00	12,820,958	70
Liabilities	5,090,396	7,341,711	1.27	1.54	27,228,391	70
Receivables	484,075	1,226,168	19.33	4.30	6,757,000	70
Net income (loss)	632,707	1,180,884	9.15	2.82	6,122,301	70
Revenue	18,064,059	23,895,301	1.09	1.42	90,826,170	70

the company's ability to generate coverage for its ongoing expenses. Obviously, too, a dearth of revenue indicates potential for firm failure (McCracken & O'Connell, 2009). We divide our sample into two sets: 50% of the observations into training set and the other 50% into validation set. Approximately 10% of the companies in each set were bankrupt companies.

One important way of judging the performance for any classification procedure is to calculate its error rates or misclassification probabilities. The performance of a sample classification function can be evaluated by calculating the actual error rate (AER). The AER indicates how the sample classification function will perform in future samples. Just as with the optimal error rate, it cannot be calculated because it depends on an unknown density function. However, an estimate of a quantity related to the AER can be calculated.

There is a measure of performance that does not depend on the form of the parent population, which can be calculated for any classification procedure. This measure is called the APparent Error Rate (APER). It is defined as the fraction of observations in the training sample that are misclassified by the sample classification function.

The APER can be easily calculated from the confusion matrix, which shows actual versus predicted group membership.

For n_1 observations from Π_1, n_2 observations from Π_2, the confusion matrix is given in Table 2 (Morrison, 1969):

The apparent error rate is thus APER $= 1 - (n_{11} + n_{22}/n)$ or, in other words, the proportion of items in the training set that are misclassified, where, $n = n_1 + n_2$.

The APER is intuitively appealing and easy to calculate. Unfortunately it tends to underestimate the AER, and the problem does not appear unless the sample sizes of n_1 and n_2 are very large. This very optimistic estimate occurs because the data used to build the classification are used to evaluate it.

Table 2. Example of Confusion Matrix.

		Predicted Membership		
		Π_1	Π_2	
Actual membership	Π_1	n_{11}	n_{12}	n_1
	Π_2	n_{21}	n_{22}	n_2

Note: n_{11} = number of Π_1 items correctly classified as Π_1 items.n_{12} = number of Π_1 items misclassified as Π_2 items.n_{21} = number of Π_2 items misclassified as Π_1 items.n_{22} = number of Π_2 items correctly classified as Π_2 items.

Table 3. Training Set.

		Predicted Membership		
		Bankrupt	Non-bankrupt	Percent correct
Actual membership	Bankrupt	3	0	100%
	Non-bankrupt	0	32	100%

Table 4. Validation Set.

		Predicted Membership		
		Bankrupt	Non-bankrupt	Percent correct
Actual membership	Bankrupt	2	1	66%
	Non-bankrupt	0	32	100%

The error rate estimates can be constructed so that they are better than the apparent error rate. They are easy to calculate, and they do not require distributional assumptions. Another evaluation procedure is to split the total sample into a training sample and a validation sample. The training sample is used to construct the classification function and the validation sample is used to evaluate it. The error rate is determined by the proportion misclassified in the validation sample. This method overcomes the bias problem by not using the same data to both build and judge the classification function.

Table 3 shows the result for the training set. The training set successfully classifies the bankrupt firms giving 100% classification accuracy. However, the strength of a discriminant model is judged by its performance on the validation set. Table 4 shows the results of the validation set. On the validation set, our model successfully classifies approximately 67% of the bankrupt firms.

CONCLUSIONS

The model, a relatively simple one, has demonstrated the capacity to correctly predict 100% of non-bankrupt retail stores and 66% of bankrupt retail firms. It stands in contrast to much more complex models, such as that of Sun (2007), which involve much higher costs of data compilation and manipulation. Sun's (2007) hazard model, using a cross-industry sample of bankrupt and non-bankrupt firms, correctly classified 82.7% of non-bankrupt firms, and 84% of bankrupt firms at a cost ratio of 100:1. In Sun's study, the complex hazard model outperformed her "audit" model, which correctly predicted 96.6% of non-bankrupt firm membership but only 44% of bankrupt firm membership. Sun's less "costly" models performed even more poorly.

This chapter demonstrated that a relatively simple model and prediction procedure can predict bankruptcy status using easily available and understood accounting variables. By limiting our study to retail firms alone, we avoided problems arising from differences in accounting method use across industries. Further research should address extension of the relatively simple method used here to develop models for specific use in other industries that may not be heavily impacted by inventory stores or receivable balances, as is the case with retail firms.

NOTES

1. The Merton model states that "a firm's equity value can be viewed as a European call option on a firm's assets" (Ghargori et al., 2006, p. 210). The authors explain that European call options are exercisable only at maturity.

2. Sun (2007) found that for the period of her study, 1991–2002, 3.28% of general merchandise stores entered bankruptcy, as did 2.32% of miscellaneous retail firms. On average, across all industries Sun surveyed, 1.35% of firms entered bankruptcy during her study's period.

REFERENCES

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance*, 589–609.

Anderson, J., & Narasimhan, R. (1979). Assessing project implementation risk: A methodological approach. *Management Science*, *25*(6), 512–521.

Awh, R. Y., & Waters, D. (1974). A discriminant analysis of economic, demographic, and attitudinal characteristics of bank charge-card holders: A case study. *Journal of Finance*, *29*, 973–983.

Capon, N. (1982). Credit scoring systems: A critical analysis. *Journal of Marketing*, *46*, 82–91.

Dillon, W., Calantone, R., & Worthing, P. (1979). The new product problem: An approach for investigating product failures. *Management Science*, *25*(12), 1184–1196.

Erenguc, S., & Koehler, G. (1990). Linear programming methods for discriminant analysis: Introduction. *Managerial and Decision Economics*, *11*(4), 213–214.

Fisher, R. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, *7*, 179–188.

Gallant, S. (1988). Connectionist expert systems. *Communications of the ACM*, *31*(2), 152–169.

Ghargori, P., Chan, H., & Faff, R. (2006). Investigating the performance of alternative default-risk models: Option based versus accounting-based approaches. *Australian Journal of Management*, *31*(2), 207–234.

Ketzner, J. (2005). Weathering the storm of unforeseeable bad-debt losses. *Business Credit*, *2005*(5), 18–20.

McCracken, J., & O'Connell, V. (2009). Wave of bankruptcy filings expected from retailers in wake of holidays. *Wall Street Journal (Eastern Edition)*, New York, January 12, p. B.1.

Meyers, L., Gamst, G., & Guarino, A. (2005). *Applied multivariate research: Design and interpretation*. Thousand Oaks, CA: Sage Publications, Inc.

Morrison, D. (1969). On the interpretation of discriminant analysis. *Journal of Marketing Research*, *6*, 156–163.

Myers, J., & Forgy, E. (1963). The development of numerical credit evaluation systems. *Journal of the American Statistical Association*, *58*(303), 799–806.

Ooghe, H., & De Prijcker, S. (2008). Failure processes and causes of company bankruptcy: A typology. *Management Decision*, *46*(2), 223–242.

Parnes, D. (2009). The systematic and idiosyncratic modules of bankruptcy risk. *The Journal of Credit Risk*, *5*(1), 25–46.

Robertson, T. S., & Kennedy, J. N. (1968). Prediction of consumer innovators-multiple discriminant analysis. *Journal of Marketing Research*, *5*, 64–69.

Shmueli, G., Patel, N., & Bruce, P. (2006). *Data mining for business intelligence: Concepts, techniques, and applications in Microsoft Office Excel with XLMiner*. New Jersey: Wiley.

Sun, L. (2007). A re-evaluation of auditors' opinions versus statistical models in bankruptcy prediction. *Review of Quantitative Finance and Accounting*, *28*, 55–78.

Tam, K., & Kiang, M. (1992). Managerial applications of neural networks: The case of bank failure predictions. *Management Science*, *38*(7), 926–947.

Tansey, R., White, M., & Long, R. (1996). A comparison of loglinear modeling and logistic regression in management research. *Journal of Management*, *22*(2), 339–358.

Tunick, B. (2002). Kmart's bankruptcy snowball: Analysts cried sell-and didn't even know about surety bond problem. *The Investment Dealers' Digest*, New York, January 28, pp. 6–7.

Westin, R. (1973). *Predictions from binary choice models*. Discussion paper no. 37. Northwestern University, Evanston, IL, USA.

Zahavi, J., & Levin, N. (1997). Issues and problems in applying neural computing to target marketing. *Journal of Direct Marketing*, *11*(4), 63–75.

A MULTI-CRITERIA DECISION MODEL FOR FIXED INCOME SECTOR ALLOCATION FOR ENDOWMENT FUNDS

Karen M. Hogan, Amy F. Lipton and Gerard T. Olson

ABSTRACT

Bond investing requires decision-making on multiple levels. Some criteria are qualitative, some are quantitative, and there may be conflicting objectives such as avoidance of credit risk versus need for income. Since managers of endowment funds must allocate their assets based on numerous dimensions, a multi-criteria decision model can help to evaluate competing criteria. We describe the Analytical Hierarchy Process (AHP), which allows investors to integrate multiple decision criteria, and apply the model to the sector allocation problem faced by managers of endowment portfolios. The AHP gives rise to a flexible model for bond investors for a range of economic scenarios, risk profiles, and time horizons.

Financial Modeling Applications and Data Envelopment Applications
Applications of Management Science, Volume 13, 71–85

ISSN: 0276-8976/doi:10.1108/S0276-8976(2009)0000013007

1. INTRODUCTION

The recent economic contraction and stock market decline of 40% on broad market indices have caused a significant problem for managers of endowment funds. In particular, the losses have severely impacted endowment funds' ability to meet current income needs. University administrators face difficult choices concerning staffing, expansion, and financial aid. Non-profit organizations have been forced to cut back significantly on the services they provide directly to their constituencies.

The US market has $30 trillion outstanding in bond investments, including US Treasury, corporate, mortgage-backed, and municipal debt (Securities Industry and Financial Market Association, 2008)). In a survey of colleges and universities encompassing $340 billion in endowment funds, 12% of assets were allocated to fixed income in 2007, 73% of which was actively managed (Commonfund Institute, 2008a). Foundations representing $195 billion in total funds allocated 15% of assets to fixed income in 2007, 85% of which was actively managed (Commonfund Institute, 2008b). In both groups, the allocation to fixed income is greater, the smaller the endowment, ranging from 11% for educational institutions over $1 billion to 26% for schools under $10 million (Commonfund Institute, 2008a) and from 14% for foundations over $1 billion to 19% for those between $51 and $100 million (Commonfund Institute, 2008b). Given the importance of fixed income in generating cash flow to fund the needs of these organizations, endowment managers need a framework to optimally allocate within the bond sector.

Once the manager of the endowment makes the asset allocation decision, the importance of the fixed income portion would be on the short- to intermediate-term income generation. However, bond managers of endowments still need to be concerned with maintaining principle and providing some long-term capital gain.

The purpose of this chapter is to use a multi-criteria approach for designing the sector allocation within the fixed income asset class. Although investment policy and asset allocation are primarily important to portfolio performance, Brinson, Singer, and Beebower (1991) point out that selection within asset classes can also contribute to portfolio performance. When designing a portfolio, investors must decide what mix of securities to use to satisfy their need for capital appreciation (growth) versus income. For example, retirement funds have a long-term, capital appreciation focus while endowments must balance capital appreciation with a short- to intermediate-term income need.

There are also other variables to consider when making the sector allocation decision. One aspect of this decision concerns the amount of risk the endowment is willing to take. Each sector of the bond market has different characteristics with respect to credit exposure, interest rate risk, and cash flow structure, which the endowment must integrate into the decision making process. Another consideration when designing a portfolio is the degree of liquidity the investor wishes to maintain. Liquid investments can be converted to cash without substantial loss of value in a relatively short period of time. On one hand, the bond market is less liquid, particularly for smaller endowments; on the other hand, bonds can provide considerably more current income than stocks. Even within the bond sector, investment horizons can differ: endowment funds have relatively short- to intermediate-term needs, but also consider capital preservation and appreciation. Fixed income managers can use different types of bonds to design portfolios with short-term, intermediate-term, or long-term horizons.

Despite the size of the fixed income investment universe, and its varied investor base, there is relatively little academic research on allocation decisions among bond sectors. In addition, there are no models that incorporate both the qualitative and the quantitative variables to allow the investor to make optimal decisions. Herold, Maurer, and Purschaker (2005) develop a strategy for dynamically allocating against a fixed income benchmark portfolio based on determining an investor's tolerance for a shortfall and budgeting that risk. This approach takes the benchmark portfolio as given and does not discuss how the investor might determine the appropriate benchmark. Herold (2003) constructs a model for incorporating a fixed income manager's qualitative views into a portfolio at the broadest level. This approach quantifies the manager's confidence in his views but does not relate them to the investor's preferences or the relative impact of the views on different bond sectors. Often the focus is on modeling the broad credit decision within one fixed income asset class such as corporate bonds (Dynkin, Hyman, & Phelps, 2004; Korn & Kovilyanskaya, 2007). Ideally, a multi-criteria process can help endowments and their bond managers integrate quantitative and qualitative factors to choose appropriate benchmarks and allocate investments within the asset class. The purpose of this chapter is to aid the endowment fund manager by developing a decision model that applies the Analytical Hierarchy Process (AHP). This model can be applied by either an internal manager or an external portfolio manager.

The AHP is a model that can integrate quantitative and qualitative information to improve bond sector allocation decisions. The model has several advantages. AHP can impose consistency and continuity to a judgmental process such as asset allocation. AHP is a multi-criteria decision support system developed by Saaty (1982), which allows a decision maker to structure a complex problem in the form of a hierarchy. The first level of the hierarchy is the goal. For this problem, the goal is to allocate assets within the fixed income sector efficiently. The next levels of the hierarchy include criteria in order of importance in achieving the goal. For this problem, an advantage of AHP is that it incorporates the systematic handling of quantitative and qualitative criteria associated with the endowment's planning horizon, investment objectives, and risk/return preferences, coupled with the manager's forecasts of economic activity, into the decision-making process. Depending on the problem, a set of sub-criteria for a given level of the hierarchy may be required. For example, a set of sub-criteria must be determined to evaluate the risk/return preferences of the endowment. The final level of the hierarchy includes the alternatives to be evaluated. For this problem, these represent the subclasses of the US domestic bond market: US Treasuries, corporate bonds, US Agency mortgage-backed securities, and money market instruments.

The AHP model has been recently applied to a multitude of different corporate and non-corporate problems to improve decision making (see, for example, Hogan, Olson, & Sillup, 2006; Hogan, Olson, & Rahmlow, 2000; Hogan & Olson, 1999; Liberatore & Nydick, 1990; Liberatore, Monahan, & Stout, 1993). It has also been used in an asset allocation context, for investment portfolios in general (Khaksari, Kamath, & Grieves, 1989), and for individual investors (Hogan & Olson, 2004; Kumar, Banu, & Nayagam, 2008). The major advantage of the model is its ability to accommodate complex qualitative and quantitative information into the decision-making process. Other advantages include its simplicity to use and its ability to apply consistency to the decision maker's judgments.

2. THE SECTOR ALLOCATION DECISION AND THE ANALYTICAL HIERARCHY PROCESS

The goal of the sector allocation decision process is to evaluate different bond asset classes in a manner that balances the endowment's planning horizon, investment objective, and risk/return preferences with the manager's forecasts and the sector characteristics. To implement the decision-making process that will allocate assets across a specified set of bonds requires two separate activities. First, the manager must identify the factors that affect the choice of bond sectors under investigation. Second, the manager must evaluate these factors to determine what percent of the endowment's wealth is allocated to each bond sector.

There are four general steps required to implement the AHP. First, the decision maker identifies the levels with their criteria and determines their relative importance in achieving the goal. Second, the decision maker determines the relative importance for each of the criteria within a given level. Pairwise comparisons must be made to determine the relative importance of the criteria in achieving the goal. Although there are many scales that can be used to compare the criteria, Saaty (1980) recommends a scale from 1 to 9 where 1 refers to "equally important," 3 "moderately more important," 5 "strongly more important," 7 "very strongly more important," and 9 "extremely more important." If more discrimination is necessary, intermediate values, 2, 4, 6, and 8, can be used.

The results of the comparisons are represented in a pairwise comparison matrix similar to Eq. (1).

$$\mathrm{W} = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1n} \\ w_{21} & \dots & \dots & w_{2n} \\ \dots & \dots & \dots & \dots \\ w_{n1} & w_{n2} & \dots & w_{nn} \end{bmatrix} = \begin{bmatrix} 1 & w_{12} & \dots & w_{1n} \\ 1/w_{12} & 1 & \dots & w_{2n} \\ \dots & \dots & 1 & \dots \\ 1/w_{1n} & 1/w_{2n} & \dots & 1 \end{bmatrix} \tag{1}$$

where w_{ij} = the relative importance of criteria i compared to criteria j; $w_{ij} = 1\ \forall\ i = j$; and $w_{ji} = 1/w_{ij}\ \forall\ i \neq j$.

If $n = 5$, W will be a 5×5 matrix with 1s along the main diagonal depicting comparison of the criteria with itself. Below the main diagonal are the reciprocals of the corresponding comparisons above the diagonal. Thus, if $n = 5$, a total of 10 comparisons must be made. In general, if there are n criteria to be compared, a total of $n(n-1)/2$ comparisons are required.

The third step in implementing AHP requires using the input comparison matrixes from step 2 to generate priority vectors at each level of the hierarchy. Finally, the priority vectors are synthesized to compute the relative contribution of the alternatives in achieving the goal.

3. AN APPLICATION OF AHP TO THE FIXED INCOME ALLOCATION DECISION

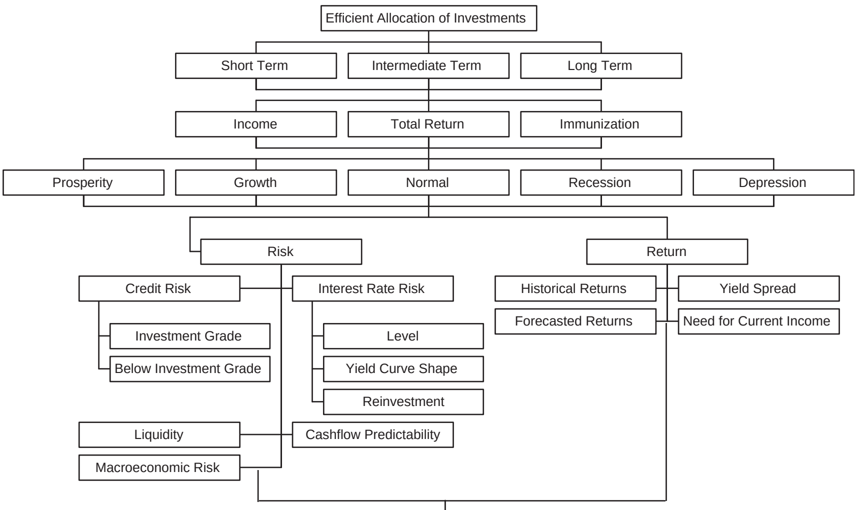
Suppose a fixed income manager is allocating among sectors of the bond market for an endowment fund. The overall goal is to determine an efficient allocation of bond investments. The manager identifies the following levels as important factors in achieving the goal: planning time horizon,

Exhibit 1. Analytical Hierarchy Process Model for Portfolio Allocation for the Fixed Income Investor.

Goal: Determine an Efficient Allocation within the Fixed Income Sector

Levels

- Planning time horizon
- Investment objective
- Economic conditions
- Risk/return preferences
  - Risk/return sub-criteria
- Bond sector allocations



- US Treasuries
- Investment Grade Corporate Bonds
- Below Investment Grade Corporate Bonds
- Agency Mortgage-Backed Securities
- Money Market Securities

investment objective, economic conditions, and risk/return preferences. Once the levels have been determined, the investor can then identify criteria and sub-criteria related to each of the levels (Exhibit 1).

The criteria for each of the levels can be described as follows:

Planning Time Horizon. The endowment has three planning horizons for investments: short term, intermediate term, and long term, where

T1 Short term, defined as less than one year.
T2 Intermediate term, defined as greater than one year and up to 10 years.
T3 Long term, defined as greater than 10 years.

Investment Objective. The endowment has stated its investment priorities: income, total return, and immunization, where

O1 Income, defined as the need for the fixed income portfolio to generate current income.
O2 Total return, defined as the need for the portfolio to provide the greatest possible combination of capital gains and income.
O3 Immunization, defined as the need for an investor to ensure, as closely as possible, a particular payout at a specific date in the future.

Economic Conditions. The manager has identified five possible states of economic activity: prosperity, growth, normal, recession, and depression, where

E1 Prosperity, defined as a significant expansion in economic activity.
E2 Growth, defined as a greater than normal expansion in economic activity.
E3 Normal, defined as a typical expanding economy.
E4 Recession, defined as a contraction in economic activity.
E5 Depression, defined as a significant contraction in economic activity.

Risk/Return Preferences. The endowment is identified as risk averse yet wishes to take advantage of opportunities to achieve significant returns, where

R1 Risk, defined as the chance, or probability, of an unfavorable event occurring to the value of the investments.
R2 Return, defined as the factors impacting an increase in the wealth position of the investor.

Risk Sub-Criteria. The manager identifies five risk factors affecting bond investments: credit risk, interest rate risk, liquidity risk, cashflow predictability, and macroeconomic shocks, where

R11 Credit risk, defined as the client's tolerance for bearing the risk of default in the investments. Credit risk is further subdivided as:
  R111 Investment grade, to what extent the endowment is willing to take credit risk within the investment grade credit ratings, defined as AAA/AAA to BBB/Baa by Standard and Poor's and Moody's, respectively, and

R112 Below investment grade, to what extent the endowment is willing to take credit risk within the below investment grade credit ratings, defined as below BB/Ba by Standard and Poor's and Moody's, respectively. It should be noted that some endowment funds may be prohibited from investing in below investment grade securities.

R12 Interest rate risk, defined as the investor's tolerance for bearing the risk of changes in interest rates. Interest rate risk is particularly critical for bond investors. Interest rate risk is further subdivided as:

R121 Level risk, defined as the change in yield-to-maturity across all bond maturities.

R122 Yield curve risk, defined as the change in the slope of the yield curve. The yield curve is the relationship between the number of years to maturity of a bond and its yield-to-maturity, or required return. When yields change more for securities of longer maturity than for bonds of shorter maturity (or vice versa), the slope of the yield curve changes.

R123 Reinvestment risk, defined as the change in the interest rate at which cashflows are reinvested within the bond portfolio.

R13 Liquidity risk, defined as the chance the endowment will be unable to convert the investment at its market value. Liquidity risk may be assessed by the frequency with which a class of bonds trades, the bid-asked spread for a class of bonds, and the loss associated with selling a particular type of bond during depressed markets.

R14 Cashflow predictability, defined as the chance the endowment will receive payments from the investment before or after they are expected. Assuming no adverse credit event, straight corporate debt has highly predictable cashflows (scheduled coupon payments and final principle payment at maturity). Callable bonds, however, may be redeemed early at the issuer's option, leading to the investor receiving principle prior to the stated maturity of the bond. Similarly, mortgage-backed securities are subject to prepayment risk, where the investor may receive principle payments earlier or later than expected, depending on homeowners' refinancing decisions. Cashflow predictability can have a significant impact on reinvestment, since investors tend to receive principle unexpectedly in low interest rate environments.

R15 Macroeconomic shocks, defined as the chance of an unexpected event occurring, such as 9/11, an unexpected adverse change in monetary policy, an unexpected change in economic activity, an unexpected change in regulations, etc., that would adversely affect the value of the investment. Changes in inflation expectations are particularly important for fixed income investors.

Return Sub-Criteria. The manager identifies four factors affecting the return of the investments: historical returns, forecasted returns, yield spread, and income, where

R21 Historical returns, defined as the returns each bond class (US Treasuries, corporate bonds, and mortgage-backed securities), has earned in the past. Historical returns can be assessed by the investor's perception of whether historical returns can be used as good or poor predictors of future returns.
R22 Forecasted returns, defined as analysts' forecasts for future returns on the bond classes. Forecasted returns can be assessed by the investor's perception of whether forecasted returns can be used as good or poor predictors of future returns.
R23 Yield spreads, defined as the risk premia over US Treasury yields attributable to credit risk, predictability of cashflows, and liquidity risk.
R24 Income, defined as the investment's ability to provide current income to the investor.

Once the criteria and sub-criteria have been identified, the next step in the process is to construct pairwise comparison matrices. In Exhibit 2 a pairwise

Exhibit 2. Pairwise Comparison Matrix for Planning Time Horizon and Computation of Local Priorities.

Pairwise Comparison Matrix

	Short Term	Intermediate Term	Long Term
Short term	1.00	3.00	7.00
Intermediate term	0.33	1.00	9.00
Long term	0.14	0.11	1.00
Column total	1.48	4.11	17.00

Adjusted Matrix and Priority Vector

	Short Term	Intermediate Term	Long Term	Priority Vector
Short term	0.68	0.73	0.41	0.606
Intermediate term	0.23	0.24	0.53	0.333
Long term	0.10	0.03	0.06	0.061

comparison matrix is presented for the planning time horizon of an endowment fund. If the investor believes that the short-term planning horizon is "moderately more important" than the intermediate-term planning horizon, a value of 3 is placed in cell w_{12} and a value of 1/3 is placed in cell w_{21}. If the investor believes that the short-term planning horizon is "strongly more important" than the long-term planning horizon, a value of 7 is placed in cell w_{13} and 1/7 is placed in cell w_{31}. If the investor believes that intermediate-term planning horizon is "extremely more important" than the long-term planning horizon, a value of 9 is placed in cell w_{23} and 1/9 is placed in cell w_{32}.

In Exhibit 2, the pairwise comparison matrix is used to estimate a priority vector that represents the relative criteria weights.[1] To determine the priority vector for the planning horizon, each column of the pairwise comparison matrix is summed and each cell is divided by its column total. The result is an adjusted matrix. The priority vector is calculated by computing the average of the entries in each row of the adjusted matrix. In this example, the priority vector for planning horizon lists a weight of .606 for short-term planning horizon, .333 for intermediate-term planning horizon, and .061 for long-term planning horizon. The results can be interpreted as the short-term planning horizon being 1.82 times more important than intermediate-term planning horizon (.606/.333), and 9.93 times more important than the long-term planning horizon conditions (.606/.061), and the intermediate-term planning horizon being 5.46 times more important than long-term planning horizon (.333/.061).

Panel A of Exhibit 3 presents priority matrixes, local weights, and global weights for a hypothetical fixed income manager for the planning time horizon, investment objective, economic conditions, and risk/return preferences levels for her endowment fund client. The local weights, or priority vectors, for each level are determined using pairwise comparisons as described earlier. For example, suppose the hypothetical endowment fund using the prescribed pairwise comparison procedure described earlier assigns local priorities for investment objective for a given short-term planning horizon as .711, .237, and .052 for income, total return, and immunization, respectively. Similarly, the investor assigns .685, .263, and .052 for the intermediate-term planning horizon and .632, .316, and .052 for the long-term planning horizon.

The global weights for one level are determined by multiplying the local priorities from the current level by the corresponding priority matrix from the level above. For example, the global weights for the investment objective

Exhibit 3. Priority Matrixes and Global Weights.

Panel A

Levels		Priority Matrixes					Local Weights	Global Weights
Planning time horizon								
T1	Short term						.606	.606
T2	Intermediate term						.333	.333
T3	Long term						.061	.061
Investment objective			T1	T2	T3			
O1	Income		.711	.685	.632			.698
O2	Total return		.237	.263	.316			.250
O3	Immunization		.052	.052	.052			.052
Economic conditions			O1	O2	O3			
E1	Prosperity		.111	.109	.094			.109
E2	Growth		.285	.315	.447			.301
E3	Normal		.285	.282	.233			.281
E4	Recession		.261	.247	.189			.254
E5	Depression		.059	.047	.038			.055
Risk/return preferences		E1	E2	E3	E4	E5		
R1	Risk	.167	.250	.750	.833	.875		.564
R2	Return	.833	.750	.250	.167	.125		.436
Risk/return sub-criteria			R1					
R11	Credit		.261					.147
R12	Interest Rate		.487					.275
R13	Liquidity		.057					.032
R14	Cashflow predictability		.096					.054
R15	Macroeconomic shocks		.099					.056
Credit risk sub-criteria			R11					
R111	Investment grade		.900					.235
R112	Below investment grade		.100					.026
Interest rate risk sub-criteria			R12					
R121	Level		.669					.326
R122	Yield curve		.088					.043
R123	Reinvestment		.243					.118

Exhibit 3. (*Continued*)

Panel A

Levels		Priority Matrixes	Local Weights	Global Weights
		R2		
R21	Historical returns	.081		.035
R22	Forecasted returns	.075		.033
R23	Yield spread	.404		.176
R24	Income	.440		.192

Panel B

Priority matrix

		R111	R112	R121	R122	R123	R13	R14	R15	R21	R22	R23	R24	R25
S1	US Treasury	.402	.195	.414	.164	.165	.383	.044	.331	.463	.065	.040	.088	.402
S2	Investment Grade	.191	.360	.183	.139	.285	.174	.205	.342	.253	.364	.160	.191	.191
S3	Below Investment Grade	.025	.032	.106	.077	.444	.033	.193	.025	.166	.113	.497	.466	.025
S4	Agency MBS	.216	.230	.266	.572	.070	.109	.520	.099	.074	.283	.226	.210	.216
S5	Money Market	.166	.182	.032	.047	.036	.300	.037	.204	.044	.174	.076	.046	.166

level depicted in Panel A of Exhibit 3 are computed as follows.

$$\begin{bmatrix} .711 & .685 & .632 \\ .237 & .263 & .316 \\ .052 & .052 & .052 \end{bmatrix} * \begin{bmatrix} .606 \\ .333 \\ .061 \end{bmatrix} = \begin{bmatrix} .698 \\ .250 \\ .052 \end{bmatrix} \tag{2}$$

The bond sectors under consideration for allocation of the investments include US Treasuries, corporate bonds (both investment and below investment grade), US agency mortgage-backed securities, and money market securities. Panel B of Exhibit 3 lists the priority matrix for five bond sectors for given risk/return sub-criteria. Multiplying the priority matrix from Panel B of Exhibit 3 by the global weights for the risk/return sub-criteria from Panel A of Exhibit 3 results in the portfolio allocations for the hypothetical endowment fund manager depicted in Exhibit 4. Given the information from the planning time horizon, investment objective, economic conditions, risk/return preferences, risk sub-criteria, and return sub-criteria levels, the endowment should invest 22.3% in Treasuries, 20.8%

Exhibit 4. Fixed Income Allocations for an Endowment Fund.

		Sector Allocation
S1	US Treasuries	.223
S2	Investment Grade Corporates	.208
S3	Below Investment Grade Corporates	.254
S4	Agency Mortgage-Backed Securities	.229
S6	Money Market Securities	.086

in investment grade corporate bonds, 25.4% in below investment grade corporate bonds, 22.9% in agency mortgage-backed securities, and 8.6% in money market instruments. This allocation would be consistent with an endowment fund's need to balance short- and intermediate-term need for income and low risk with its need for some long-term capital appreciation.

4. MODEL VALIDATION

In the aforementioned example, the objectives and preferences of the endowment fund fixed income investor are used to conduct pairwise comparisons. Since the resulting bond allocations are dependent on these comparisons, it is desirable to use the optimal set of criteria and their relative importance. Unfortunately, finance theory has not yet identified the optimal set of criteria that can be used to evaluate portfolio allocations for a bond investor with multiple planning horizons and under different economic conditions.

AHP is a multi-criteria decision support system that can integrate quantitative and qualitative information. The value of the model is dependent on the inputs of the user. Srinivasan and Kim (1988) note that it is possible that the expert's knowledge may be incorrect, and AHP can merely serve to institutionalize incorrect knowledge. Until an optimal set of criteria can be identified in the finance literature, care should be taken in the selection of the relevant inputs in determining sector allocation.

By using past investment data, endowment fund managers can use the model as a focal point for rethinking the tradeoffs among different sets of criteria. The simplicity of implementing AHP permits the manager to easily revise the criteria based on changing endowment parameters. As the endowment's needs change, the criteria and priority vector weights can be altered to reflect a new allocation of investments among bond sectors.

5. SUMMARY AND CONCLUSIONS

The effective allocation of assets within the fixed income asset class involves both subjective and objective information. A difficulty arises in the implementation of the process due to multiple evaluative criteria that may be troublesome to measure. AHP is a decision support system that can integrate both subjective and objective information to improve the efficiency of a fixed income manager's allocation of bond investments.

AHP requires the structuring of the problem into the form of a hierarchy, which consists of a goal, evaluation criteria, and possibly sub-criteria, and alternatives. Pairwise comparisons are made on items on each level of the hierarchy to the level above it and the relative importance of the items can be determined. An overall allocation for each bond sector is computed based on the endowment's specific parameters.

In this chapter, we describe AHP and apply the model to the allocation of assets among bond investments for a hypothetical endowment fund. The result is a flexible and consistent model that can reduce the risk associated with investing with multiple criteria. AHP is flexible in that the criteria and sub-criteria can be revised based on the needs of the user. Also, the relative importance of the criteria and sub-criteria can be easily recomputed with the use of a spreadsheet or dedicated software. The result is a consistent measurement scale that can be used to efficiently allocate investments. AHP for sector allocation is useful for professional managers of endowments, pensions, mutual funds, and can also be used by individuals.

NOTE

1. This procedure is a good approximation of the weights. Saaty (1980, 1982) determined the exact relative priorities for each of the *n* criteria by computing the normalized eigenvector of the maximum eigenvalue of the comparison matrix. The normalized eigenvector is computed by raising the comparison matrix to successive powers until convergence is achieved and then normalizing the results.

REFERENCES

Brinson, G., Singer, B., & Beebower, G. (1991). Determinants of portfolio performance II: An update. *Financial Analysts Journal* (May–June), 40–48.

Commonfund Institute. (2008a) *Commonfund Benchmarks Study*®. Education Endowment Report.

Commonfund Institute. (2008b) *Commonfund Benchmarks Study*®. Foundations Report.

Dynkin, L., Hyman, J., & Phelps, B. D. (2004). Optimal credit allocation for buy-and-hold investors: A top-down, strategic approach. *Journal of Portfolio Management*, *30*(4), 73–91.

Herold, U. (2003). Portfolio construction with qualitative forecasts: It doesn't need to be a black box. *Journal of Portfolio Management*, *30*(1), 61–72.

Herold, U., Maurer, R., & Purschaker, N. (2005). Total return fixed-income portfolio management: A risk-based dynamic strategy. *Journal of Portfolio Management*, *31*(3), 32–43.

Hogan, K., & Olson, G. (1999). Evaluating potential acquisitions using the analytic hierarchy process. *Advances in Mathematical Programming and Financial Planning*, *5*, 3–18.

Hogan, K., & Olson, G. (2004). A multi-criteria decision model for portfolio allocation for the individual investor. *Mathematical Programming Applications of Management Science*, *11*, 3–16.

Hogan, K., Olson, G., & Rahmlow, H. (2000). A multi-criteria model for predicting corporate bankruptcy using the analytical hierarchy process. *Applications of Management Science*, *10*, 85–102.

Hogan, K., Olson, G., & Sillup, G. (2006). Identifying compounds for pharmaceutical companies' development tracks using a multi-criteria decision model. *Applications of Management Science: In Productivity, Finance, and Operations*, *12*, 89–108.

Khaksari, S., Kamath, R., & Grieves, R. (1989). A new approach to determining optimum portfolio mix. *The Journal of Portfolio Management* (Spring), 43–50.

Korn, R., & Kovilyanskaya, H. (2007). A general framework for high yield bond investment. *International Journal of Theoretical and Applied Finance*, *10*(6), 967–984.

Kumar, K., Banu, C., & Nayagam, L. (2008). Financial product preferences of Tiruchirapalli investors using analytical hierarchy process and fuzzy multi criteria decision making. *Investment Management and Financial Innovations*, *5*, 66–73.

Liberatore, M., Monahan, T., & Stout, D. (1993). Strategic capital budgeting for investments in advanced manufacturing technology. *The Journal of Financial and Strategic Decision Making*, 55–72.

Liberatore, M., & Nydick, R. (1990). An analytic hierarchy approach for evaluating product formulations. In: A. H. Bihl (Ed.), *Computer aided formulation: A manual for implementation* (pp. 179–194). New York, NY: VCH Publishing Company.

Saaty, T. L. (1980). *The analytical hierarchy process*. New York: McGraw Hill.

Saaty, T. L. (1982). *Decision making for leaders*. Belmont, CA: Wadsworth, Inc.

Securities Industry and Financial Market Association. (2008). US key statistics. Available at www.sifma.org. Retrieved on August 2009.

Srinivasan, V., & Kim, Y. H. (1988). Designing expert financial systems: A case study of corporate credit management. *Financial Management*, *17*(3), 32–44.

SECTION B
DATA ENVELOPMENT ANALYSIS (DEA) APPLICATIONS I

RECOVERING FROM DELAYS: AN ANALYSIS OF AIRPORT OPERATIONS USING DATA ENVELOPMENT ANALYSIS

Warren T. Sutton and Seungkee Baek

ABSTRACT

This chapter evaluates the operational efficiency of major airports in the United States. The airport is defined as a major point of contact in the aviation industry, and on-time operations is regarded as a core service factor. We develop a bounded data envelopment analysis (DEA) model that evaluates the punctuality of airports and proposes a three-stage approach that analyzes not only current operations performance but also efficiency changes over time. We classify airports into several classes according to Federal Aviation Authority (FAA) definitions and compare their class efficiencies through decomposed efficiency scores. We find significant differences in efficiency scores between classifications.

Financial Modeling Applications and Data Envelopment Applications
Applications of Management Science, Volume 13, 89–112

ISSN: 0276-8976/doi:10.1108/S0276-8976(2009)0000013008

1. INTRODUCTION

This study investigates airport operations in the United States and evaluates their performance using data envelopment analysis (DEA). In many sectors, financial indicators are frequently used as an effective indicator for performance measurement. Unfortunately, these financial indicators typically fail to directly measure the operational efficiency. The importance of lean operations has intensified with an increased focus on the elimination of waste as a direct contribution to increased profit. Under the slowdown of economic growth and increased competition, the efficiency of operations should be regarded as a critical factor necessary to survival in the current economy. We therefore examine the performance of airports with a focus on operational efficiency.

Since the landmark publication by Charnes, Cooper, and Rhodes (1978), DEA is now considered a major performance evaluation tool (Cooper, Seiford, & Tone, 2006). The principal unit under investigation in DEA is the decision-making unit (DMU). DEA measures the relative efficiency of a set of DMUs using mathematical programming and computes efficiency scores, benchmarking partners, and areas for improvement for each DMU. A DMU is considered efficient when it has an efficiency score of 1. An inefficient DMU has an efficiency score different than 1, and the degree of inefficiency is calculated by the distance of the DMU's efficiency score from the desired value of 1. These inefficient DMUs are given suggestions for benchmarking partners to enhance performance; these suggestions are composed of efficient DMUs, called reference units. Thus, the benefit of using DEA in airport operations can be summarized as follows; first compare the performance of airports by their efficiency scores and then make specific recommendations for areas of improvement from the benchmarking partners. Thus, we expect that DEA is the appropriate tool for accurately analyzing airport operations.

Today, most airline companies use hub and spoke networks, which are networks that have few nodes with a high node degree and many nodes with degree one. The use of these types of networks helps airlines to maximize utilization. Most major U.S. airline companies' hub airports offer transfer flights, which are flights where the hub airport is neither the origin nor the destination of the enplaned passengers. Non-hub airports are not required to offer transfer flights, and thus, a hub airport is much more likely to be crowded by flights and passengers. The efficient operation of hub airports receives higher priority in the aviation industry, leading to a possible neglect of non-hub airports in terms of efficiency. Sarkis (2000) attempted to prove that a hub airport is more efficient than non-hub airport but failed to show sufficient evidence of the existence of significant differences in the efficiency scores. A radial-based efficiency measurement was used (Sarkis, 2000), which assumes proportional change among inputs or outputs. In contrast to this approach, we use a non-radial-based efficiency measure that allows for non-proportional rates of substitution, as is the case in the aviation industry. Also, we decompose the efficiency scores into several components, pure technical efficiency, scale efficiency, and mix efficiency to perform an in-depth analysis that determines the factors that lead to efficiency differences. While Sarkis (2000) defined hub airports as airports assigned as such by airline companies, we apply the definition of the Federal Aviation Authority (FAA), which classifies hub airports into three categories (large, medium, small hub airports) according to the percentage of the total of national passengers enplaned. The FAA classification of hub airports is a more robust definition that encompasses the definition of Sarkis (2000). In general, most airports that are defined as hubs by individual airline companies are actually considered large hubs by the FAA classification. This chapter compares efficiencies among hub and non-hub airports to determine differences in the classifications.

Previous researchers in this field have indicated that the change in efficiency scores over time needs to be addressed. Gillen and Lall (1997) measured the efficiency of airport operations over five years and made a comparison of the efficiency scores per year. These studies have been used as a basis for additional research by Parker (1999), Sarkis (2000), Adler and Golany (2001), Fernandes and Pacheco (2002), Bazargan and Vasigh (2003), and Pels, Nijkamp, and Ritveld (2003). The effect of the incidents of September 11 on the airline industry is well documented, and lingering effects have been experienced by several airlines and airports even years later. An industry expert, Gordon Bethune (2005), argues for the need for smart government investment in airports to "fix" the airline industry. This impending investment opportunity makes it necessary for decision makers to identify individual airports that are in position to make a positive impact on the airline industry as a whole. Thus, a measurement tool to identify efficient operations is needed to identify and understand trends for airport efficiency. We examine changes in efficiency using a Malmquist index, which divides the cause of efficiency change into two categories; the change in efficiency due to the performance of the specific DMU and the change in efficiency due to the overall technical change. Moreover, we analyze the scale efficiency changes by using the definition of Ray and Delsi (1997) and work to clarify the factors of efficiency change that are caused by the efforts of the airport itself versus any overall technical improvement in the aviation industry.

Airports are the first point of contact for customers and a primary point for receiving service from the aviation industry. The importance of customer satisfaction should not be ignored; however, it is difficult to find research that evaluates airport performance from the customer's perspective. Yet, it is widely recognized that speed of service is the most critical evaluation factor of the aviation industry by customers (Bethune, 2005). Thus, all parts of the aviation industry, from airlines and airports to the Transportation Security Administration (TSA), should make earnest efforts to increase the timeliness of their operations. In particular, the airports themselves have an especially critical role because they control many of the operations around the on-time performance of flights. According to the Bureau of Transportation Statistics in 2006, more than half of the causes of flight delays result from the airport operations themselves. It is important to note that from a customer's perspective, on-time departures should be regarded as the major performance indicator in airport operations and that improving the efficiency of airport operations could eventually result in an increase of on-time departures as well as overall customer satisfaction (Abdelghany, Shah, Raina, & Abdelghany, 2004). Thus, we utilize a DEA model that focuses on the on-time performance of airports and employs that as a key factor to evaluate the efficiency of airport operations, which will directly enhance customer satisfaction.

The remainder of this chapter is organized as follows. Section 2 provides a review of previous research regarding the analysis of airport operations using DEA. Section 3 describes the approach and development of the DEA model. An analysis and collection of a four-year dataset of major U.S. airports is highlighted in Section 4. Next, we discuss managerial and policy implications in Section 5 and finally provide conclusions and propose possible directions for future research in Section 6.

2. AIRPORT PERFORMANCE

2.1. Operation Process

Gillen and Lall (1997) separate airport operations into two areas: terminal service and movement. The terminal service controls passenger movement, while the movement relates to flight operations such as takeoffs and landings. However, a large portion of terminal service is run by the individual airline companies, since they have the responsibility to provide safe and comfortable transportation services to their own passengers.

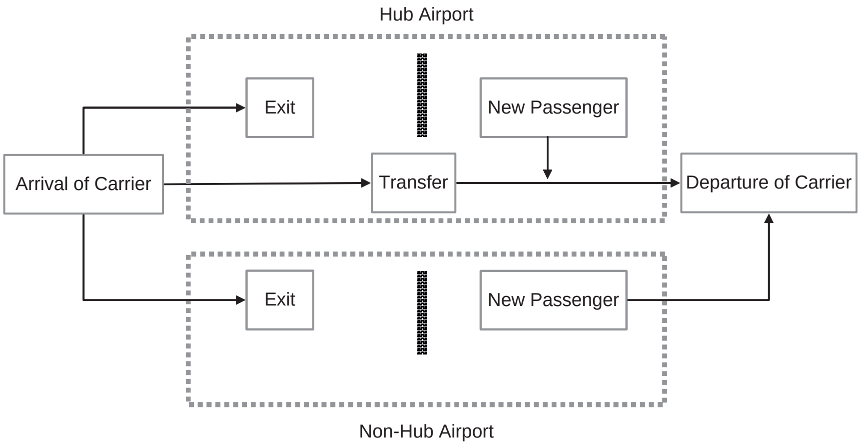


Fig. 1. Map of Airport Service Process.

We can reason that the role of the airport remains to manage the physical structures such as the gates and convenience facilities, while individual airlines and other agencies control the flow of passengers. In this study, these two operations can be considered as a single process. Although the FAA does not include transfer flights in their definition of hub airports, we assume that the hub airport can provide transfer flights while non-hub airports generally do not. Fig. 1 shows this definition of the airport service operation process.

We regard on-time departure as a core function of an airport that obtains customer satisfaction that is consistent with Abdelghany et al. (2004) who mention that customer satisfaction is the "key factor" in both maintaining current and bringing in new customers. While delayed arrival and extreme weather conditions can cause fluctuations in on-time performance, we consider those components to be uncontrollable environmental factors. Thus, the primary objective of airport operations in this study is to increase the on-time departure rate.

2.2. Previous Research

Table 1 reports typical input/output structures of selected previous research.

From Table 1, most previous research uses fixed assets as input and financial indicators as output. Therefore, productivity measured can be interpreted as the utilization rate of fixed assets over the revenue of the airport.

Table 1. Summary of Previous Research.

Research	DMU	Data Period
	Input Output DEA model	
Gillen and Lall (1997)	21 of the top 30 airports in the United States	1989–1993
	Terminal: no. of runways, no. of gates, terminal area, no. of employees, no. of baggage collection belts, and no. of public parking spots Movements: airport area, $ of runways, runway area ,and no. of employees	
	Terminal: no. of passenger and pounds of cargo Movements: air carrier movements, commuter movements	
	Terminal : BCC-DEA Movement : CCR-DEA	
Sarkis (2000)	44 of the top 80 U.S. airports	1990–1994
	Operating cost, no. of employees, no. of gates, and no. of runways	
	Operating revenues, no. of aircraft movements, general aviation, total passengers, total freight	
	CCR-DEA and BCC-DEA	
Bazargan and Vasigh (2003)	15 small, medium, and large U.S. hub airports	1996–2000
	Operating expenses, non-operating expenses, no. of runways, and no. of gates	
	No. of passengers, no. of air carrier operations, no. of other operations, aeronautical revenue, non-aeronautical revenue and percentage of on-time operations	
	CCR-DEA	
Pels et al. (2003)	33 European airports	1995–1997
	ATM: Airports surface area, no. of aircraft parking position, no. of remote aircraft parking position, and no. of runways	
	APM: no. of check-in desks, no. of baggage claim units, terminal size, and no. of aircraft parking position	
	CCR-DEA, SFA	

2.3. Hub versus Non-Hub Airports

The FAA distinguishes hub and non-hub airports by the number of passenger enplaned. The larger the number of passengers enplaned, the more flights operated, and therefore, it is not illogical that some flights at larger airports are used as transfer flights. Thus, these airports are usually also considered as hubs by major airline carriers. As shown in Fig. 1, a unique function that a hub airport provides is transfer flights. Therefore, one can assume that the role of providing transfer flights is implicitly embedded into the FAA definition of a hub airport. While Sarkis (2000) implements the definition of hub airport directly from airline companies, we expand his definition by adding the three categories used by the FAA. We also hypothesize that the difference among large, medium, and small hub airports is the number of transfer flights offered. We find that most hub airports assigned by major airline companies belong to the large hub classification in the FAA definition. As the aviation industry grows and expands, it can be expected that the demand for hub airports will also increase. Thus, it would be reasonable to surmise that current medium or small hub airports would be good candidates for airlines to investigate for expansion as potential hub airports, as commonly seen in many European budget airlines. Adler and Berechman (2001) indicate that an efficiently operated airport strongly influences the airlines' choice of hub locations.

A multi-dimensional comparison of efficiency among airports is conducted comparing both radial and non-radial-based efficiency measures and verification of significant differences among classification of airports. Next, a comparison of decomposed efficiency scores is made, and the factors that lead to efficiency differences are identified. Finally, efficiency changes among airports are examined using the Malmquist index.

While comparisons between the size and scale of hub and non-hub airports cannot be made, it could be easily expected that the returns-to-scale (RTS) of large hub airports is different from small hub or non-hub airports. Thus, we use the identification of RTS presented by Seiford and Zhu (1999) to compare RTS among airports.

3. MODEL

3.1. Overview

Our approach utilizes a three-stage DEA model to evaluate airport operation; the structure is shown in Fig. 2. In the first stage, we measure the radial and

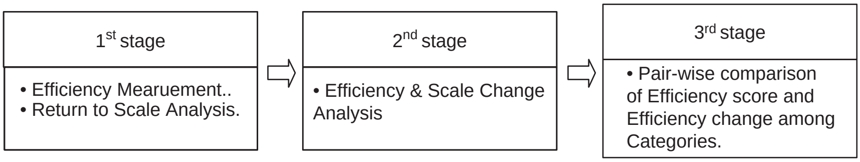


Fig. 2. Structure of Our Research.

non-radial efficiency of airports. As mentioned in the Section 2, the number of on-time departing flights is one of the focal outputs. However, it is important to note that the number of on-time departing flights cannot exceed the number of scheduled departure flights. Therefore, a bounded DEA model that applies the additional constraints of restricting the maximum number of departure flights is necessary. In the second stage, the source of efficiency change is identified using the Malmquist index. We also apply a bounded DEA model to measure catch-up (CU) and frontier shift (FS) effect. In the third stage, the differences in efficiency among airports are compared, and then, the managerial implications for the airports are analyzed.

3.2. First Stage: Efficiency Decomposition

As noted in Table 1, we reviewed the type of DEA model and input/output structures from previous research. The radial-based DEA models employed in these previous studies assume that all of the inputs or outputs can be proportionally changed, in contrast to non-radial-based DEA models.

Charnes-Cooper-Rhodes (CCR), Banker-Charnes-Cooper (BCC) and slack-based measurement (SBM) efficiency scores are measured in the first stage, and the efficiency scores are decomposed into pure technical, scale, and mix efficiency. Before evaluating DMUs, we apply additional constraints to the standard DEA model. Since customer satisfaction is taken into account in this study, on-time departures are not overlooked. We use an output-oriented approach, using on-time departure as a factor, whereas the amount of correction should not exceed the number of scheduled flights. Therefore, we add bounded constraints to the DEA models.

If we assume that there are n ($k = 1, \ldots, n$) DMUs that convert m ($i = 1, \ldots, m$) inputs into p ($j = 1, \ldots, p$) outputs, we therefore suggest an output-oriented bounded variable model to assess the precise operation of airports. Models 1(a–c) show the set of equations used to represent the Bounded CCR, Bounded BCC, and Bounded SBM models, respectively. In the following models, the variable λ is a $[nx1]$ array, s^- is a $[mx1]$ array, and

s^+ is a [$px1$] array. X is a [mxn] matrix of inputs, Y is a [pxn] matrix of outputs, x_o is a [$mx1$] array, and y_o and u_o are both [$px1$] arrays. And θ is a scalar representing the efficiency of the DMU under evaluation.

Bounded CCR

$$\begin{aligned} &\max_{\theta,\lambda} \theta \\ &\text{s.t.} \\ &X\lambda + s^- = x_o \\ &Y\lambda - s^+ = \theta \cdot y_o \\ &Y\lambda \le u_o \\ &s^+, s^-, \lambda \ge 0 \end{aligned} \tag{1a}$$

Bounded BCC

$$\begin{aligned} &\max_{\theta,\lambda} \theta \\ &\text{s.t.} \\ &X\lambda + s^- = x_o \\ &Y\lambda - s^+ = \theta \cdot y_o \\ &Y\lambda \le u_o \\ &\sum_{k-1}^{x} \lambda_k = 1 \\ &s^+, s^-, \lambda \ge 0 \end{aligned} \tag{1b}$$

Bounded SBM

$$\begin{aligned} &\min_{s^+,\lambda} \quad \frac{1}{1 + \sum_{j=1}^{p} \frac{s_j^+}{y_{jo}}} \\ &\text{s.t.} \\ &X\lambda + s^- = x_o \\ &Y\lambda - s^+ = y_o \\ &Y\lambda \le u_o \\ &s^+, s^-, \lambda \ge 0 \end{aligned} \tag{1c}$$

In Eq. (1a), we measure the efficiency of a DMU under constant RTS. We obtain an efficiency score with variable RTS by applying convexity condition $\Sigma\lambda = 1$ to form model 1b. We define the first efficiency score as

BND-CCR, and the latter as BND-BCC. While both BND-CCR and BND-BCC are radial-based efficiency scores, we evaluate a non-radial-based efficiency score from the SBM by Tone (2001). We apply the bounded constraint to SBM and define its efficiency score as BND-SBM as seen in model 1c.

We decompose BND-CCR into scale, mix, and pure technical efficiencies using Eqs. (2a–c). The scale and mixed efficiency equations presented below are the reciprocal of the input orientated counterparts that are presented in Cooper et al. (2006).

$$\text{Scale efficiency} = \frac{\text{BND\_BCC}}{\text{BND\_CCR}} \tag{2a}$$

$$\text{Mix efficiency} = \frac{\text{BND\_CCR}}{\text{BND\_SBM}} \tag{2b}$$

$$\begin{aligned}\text{Pure technical efficiency} = \text{BND\_SBM} &\times \text{scale efficiency} \\ &\times \text{mixed efficiency}\end{aligned} \tag{2c}$$

3.3. Second Stage: Malmquist Indices/Efficiency Change

While the annual changes in efficiency can be compared using the results from the first stage, the factor that causes these differences of efficiency cannot be identified. Tone (2004) discusses the various types of Malmquist indices, which measure the relative efficiency of DMUs from each different production possibility set. The Malmquist indices can be measured by two methods: inclusive and exclusive schemes. We measure the inclusive scheme of the Malmquist index by applying a bounded constraint.

We measure the Malmquist index in both Constant returns-to-scale (CRS) and Variable returns-to-scale (VRS) environments. Tone (2004) indicated that several studies have been made to examine the effect of scale change to efficiency change. We select Ray and Delsi's (1997) methodology to measure scale change effect, since it does not require the use of additional "fictitious DMUs" and ultimately requires fewer computations than Balk's (2001) method.

We then present a pairwise comparison of decomposed efficiency score by year, which provides a basic understanding of efficiency change. However, simple pairwise comparisons cannot clarify the change that results from the

DMU's own effort versus the general increase of all DMUs in the production possibility set. Thus, we further conduct pairwise comparisons of Malmquist index analysis in the second stage. Using these comparisons, we determine which hub classifications show an increase in efficiency scores between 2002 and 2005.

4. CASE

4.1. Overview

The radial and non-radial efficiency of airports in United States is measured, and these efficiencies are decomposed into pure technical, scale, and mix efficiency. We are then able to make comparisons among hub and non-hub airport based on classifications set by the FAA. The efficiency of the airports is further examined using the Malmquist index.

4.2. Data

In this section, we analyze four years (2002–2005) of data from 67 airports in United States; this data was collected from the FAA, and the input/output structure is given in Table 2 (U.S. Department of Transportation, 2005).

In Table 2, operational revenue is defined as the revenue that comes from the payment by airline companies for using landing/takeoff facilities while non-operational revenue includes all other revenues. Revenue from parking lots, restaurants, and the other convenience facilities are included in the non-operational revenue.

As discussed previously in Section 2, the on-time arrival of flights is used as an input. Using the bounded models 1(a–c), airport operations are analyzed. We use classifications set by the FAA to define hub and non-hub airports.

Table 2. Input/Output Structure.

Input	Output
No. of runway, no. of gates, no. of scheduled arrivals	Amount of operational revenue, amount of non-operational revenue, and no. of on-time departure

4.3. Result

4.3.1. First Stage

We evaluate four years of data from 67 airports. Since the yearly change in the efficiency score is not compared, we therefore regard all four years of data set as a single production possibility set. We classify the number of efficient DMUs by the type of hub. Table 3 provides the summary.

Table 3 demonstrates that most of the DMUs in small or medium hub airports show scale efficiency. We can assume that the size and scale of large hubs is so large that the scale efficiency can not be increased. We verify this argument by examining the RTS of each type of airport. Table 4 provides the summary of the distribution of RTS.

From Table 4, we find that nearly all efficient non-hub airports are increasing RTS while more than half of efficient large hub airports are decreasing RTS. It is natural that the small airports have more growth potential than a large airport, since larger airports are closer to their operational capacity.

Table 3. Summary of First Stage.

Category	Total	Number of Efficient DMUs					
		CCR	BCC	SBM	SE	ME	PTE
Non-hub (%)	20	1 (5.00)	10 (50.00)	1 (5.00)	8 (40.00)	1 (5.00)	1 (5.00)
Small hub (%)	64	4 (6.25)	4 (6.25)	4 (6.25)	55 (85.94)	4 (6.25)	4 (6.25)
Medium hub (%)	100	5 (5.00)	8 (8.00)	5 (5.00)	84 (84.00)	5 (5.00)	5 (5.00)
Large hub (%)	84	6 (7.14)	17 (20.24)	6 (7.14)	26 (30.95)	6 (7.14)	6 (7.14)
Total	268	16	39	16	173	16	16

Table 4. Summary of RTS.

Category	Total	Efficient DMUs-RTS			All DMUs-RTS		
		Increasing	Constant	Decreasing	Increasing	Constant	Decreasing
Non-hub (%)	20	9 (90.00)	1 (10.00)	0 (0.00)	12 (60.00)	8 (40.00)	0 (0.00)
Small hub (%)	64	0 (0.00)	4 (100.00)	0 (0.00)	6 (9.38)	55 (85.94)	3 (4.69)
Medium hub (%)	100	0 (0.00)	5 (62.50)	3 (37.50)	1 (1.00)	84 (84.00)	15 (15.00)
Large hub (%)	84	0 (0.00)	6 (35.29)	11 (64.71)	0 (0.00)	26 (30.95)	58 (69.05)
Total	268	9	16	14	19	173	76

4.3.2. Second Stage

In the second stage, we conduct a Malmquist index analysis, as depicted in Table 5.

When the Malmquist index is greater than 1, the DMU has a substantial increase in its productivity. From Table 5, the small hub shows consistent productivity growth within the past four years. We verify the significant difference among categories in the next stage.

Ray and Delsi (1997) suggest a methodology to identify the influence of scale change on efficiency change. We measured scale changes between years by categories, as summarized in Table 6.

As seen in Table 6, it is clear that more than half of the small and medium hubs have a value of scale change that is greater than 1, which means that the scale change has increased over time. Thus, as we found in the first stage in the RTS analysis, the small and medium hub airports have more potential for growth than large hub airports.

4.3.3. Third Stage

The objective of the third stage is to identify significant differences among airport categories. First, the radial and non-radial-based efficiency scores are compared. Table 7 reports the average efficiency scores of each airport category. We conducted a Wilcoxon Rank Sum test to identify differences in efficiency scores, as seen in Table 8.

These tables show that we are able to find significant differences between hub and non-hub airports by using a non-radial-based efficiency measurement based on our modified definition of a hub airport.

5. DISCUSSION

5.1. Efficiency Decomposition

The results of the efficiency decomposition show that there is sufficient evidence of scale efficiency existing in small and medium hubs, but not in non-hubs and large hubs, as given in Table 3. Scale efficiency is a measure of how much the efficiency score is changed when the convexity constraint $\Sigma\lambda = 1$ is included in model 1a to yield the aforementioned BND_BCC (Eq. (1b)). When the scale efficiency score is less than 1, it is an indication

Table 5. Result of Malmquist Index.

Category	Total	2002–2003			2003–2004			2004–2005			2002–2005		
		MI < 1	MI = 1	MI > 1	MI < 1	MI = 1	MI > 1	MI < 1	MI = 1	MI > 1	MI < 1	MI = 1	MI > 1
Non-hub	5	4	0	1	2	0	3	2	0	3	3	0	2
%		80.0	0.0	20.0	40.0	0.0	60.0	40.0	0.0	60.0	60.0	0.0	40.0
Small hub	16	8	0	8	6	0	10	6	1	9	9	0	7
%		50.0	0.0	50.0	37.5	0.0	62.5	37.5	6.3	56.3	56.3	0.0	43.8
Medium hub	25	15	2	8	20	1	4	7	1	17	19	2	4
%		60.0	8.0	32.0	80.0	4.0	16.0	28.0	4.0	68.0	76.0	8.0	16.0
Large hub	21	11	2	8	14	2	5	5	3	13	10	3	8
%		52.4	9.5	38.1	66.7	9.5	23.8	23.8	14.3	61.9	47.6	14.3	38.1
Total	67	38	4	25	42	3	22	20	5	42	41	5	21

Table 6. Result of Scale Change.

Category	Total	2002–2003			2003–2004			2004–2005			2002–2005		
		SC < 1	SC = 1	SC > 1	SC < 1	SC = 1	SC > 1	SC < 1	SC = 1	SC > 1	SC < 1	SC = 1	SC > 1
Non-hub	5	3	0	2	2	0	3	2	0	3	3	0	2
%		60.0	0.0	40.0	40.0	0.0	60.0	40.0	0.0	60.0	60.0	0.0	40.0
Small hub	16	3	0	13	10	0	6	5	1	10	6	0	10
%		18.8	0.0	81.3	62.5	0.0	37.5	31.3	6.3	62.5	37.5	0.0	62.5
Medium hub	25	11	2	12	10	1	14	7	1	17	13	2	10
%		44.0	8.0	48.0	40.0	4.0	56.0	28.0	4.0	68.0	52.0	8.0	40.0
Large hub	21	10	2	9	12	2	7	10	3	8	11	3	7
%		47.6	9.5	42.9	57.1	9.5	33.3	47.6	14.3	38.1	52.4	14.3	33.3
Total	67	27	4	36	34	3	30	24	5	38	33	5	29

Table 7. Descriptive Statistics of Efficiency Scores.

	CCR-O					BCC-O					SBM-O				
	N	Mean	Standard deviation	Minimum	Maximum	*N*	Mean	Standard deviation	Minimum	Maximum	*N*	Mean	Standard deviation	Minimum	Maximum
NH	20	1.208	0.095	1	1.38	20	1.111	0.139	1	1.38	20	2.615	1.462	1	6.68
SH	64	1.208	0.086	1	1.393	64	1.202	0.085	1	1.393	64	1.941	0.55	1	3.623
MH	100	1.195	0.074	1	1.316	84	1.12	0.101	1	1.395	100	1.946	1.472	1	15.411
LH	84	1.204	0.101	1	1.449	100	1.183	0.081	1	1.316	84	2.196	4.252	1	39.743

Table 8. Result of Wilcoxon Rank Sum Test for Efficiency Scores.

	LH-MH	LH-SH	LH-NH	MH-SH	MH-NH	SH-NH
CCR-O						
p-Value	0.2303	0.3457	0.5016	0.0210	0.0859	0.3135
BCC-O						
p-Value	0.0006	1.468E-05	0.5540	0.0247	0.2471	0.0333
SBM-O						
p-Value	0.0781	0.0018	0.0003	0.0479	0.0057	0.5016

that the airport under consideration benefits from the convexification of the frontier in the BND_BCC model. This leads to the significant differences that can be seen in the efficiency scores of the hub classifications. Table 9 gives the *p*-values for the Wilcoxon Rank Sum test of the hub classifications indicating significant differences at most reasonable significance levels, between all pairwise comparisons except the large and small hubs groups and the small and medium hubs groups.

This leads to the conclusion that large hubs are not able to perform at the level that would be expected of airports of that magnitude. A consequence of this is that hubs can be built too big to ever be able to achieve efficiency. On the contrary, the non-hubs also do not perform well in scale efficiency indicating that an increase in scale is necessary.

Likewise the pairing of efficiency groupings among the small and medium hubs groupings and large and non-hubs groupings continues when the pure technical efficiency is considered as evidenced in Table 9, which provides the *p*-values for the Wilcoxon Rank Sum test.

The grouping of the pure technical efficiency scores is a little surprising because it matches exactly with the results from the scale efficiency but shows that the small and medium hub groups are once again able to out perform the large and non-hub groups.

However, the results from the mixed efficiency score are quite different. In this case, the larger hubs show a clear ability to outperform the smaller hubs as evidenced in the mean ordering and *p*-values in Table 9.

The mixed efficiency score, as indicated in Eq. (2a), is an indication of the amount of inefficiency that is unaccounted for by the use of a radial model. A radial model ignores slack when calculating the efficiency score. Thus lower mixed efficiency scores imply that there is a larger amount of slack that is not included in the efficiency score given by the BND_CCR model. Our results show that this phenomenon is more prevalent as hub size decreases.

Table 9. Mean Ordering and p-Values of Wilcoxon Rank Sum Test for Scale, Pure Technical and Mixed Efficiency Scores.

Scale Efficiency			Pure Technical Efficiency			Mixed Efficiency		
p-values			*p*-Values			*p*-Values		
SH-NH	MH-NH	LH-NH	SH-NH	MH-NH	LH-NH	SH-NH	MH-NH	LH-NH
0.0019	0.0019	0.4348	0.0017	0.0010	0.2471	0.7089	0.0111	0.0003
MH-SH	LH-SH	LH-MH	MH-SH	LH-SH	LH-MH	MH-SH	LH-SH	LH-MH
0.7439	2.90E-08	3.40E-08	0.1032	0.0119	2.18E-05	0.1004	0.0007	0.0177
Mean Ordering			Mean Ordering			Mean Ordering		
Small hubs		0.9950	Non-hubs		1.3239	Large hubs		0.7669
Medium hubs		0.9902	Large hubs		1.3024	Medium hubs		0.7018
Large hubs		0.9328	Small hubs		1.2151	Small hubs		0.6664
Non-hubs		0.9196	Medium hubs		1.2082	Non-hubs		0.5686

5.2. Returns-To-Scale

The results of the RTS of the dataset indicate that there is a clear ordering among the hub classifications. As expected an increased hub size is more likely to experience decreasing RTS. Conversely, the smaller hub is more likely to experience increasing RTS. This demonstrates that non-hubs and small hubs dominate the increasing RTS portion of the technology, while the medium and large hubs are concentrated on the constant and decreasing RTS parts of the technology; this trend can be observed in Table 4.

This finding is important since it points to a key managerial implication about potential return-on-investment and capital expenditures. Traditionally, a large focus is placed on improvements in the high-volume large hubs. However, our results suggest that this strategy should not be employed when optimizing for efficiency. The non-hubs clearly show that they dominate the increasing RTS portion of the frontier and would thus yield higher return-on-investment and should be given more consideration for capital investment and improvement programs.

5.3. Hub Comparisons

The research question which was to examine the differences in efficiency of hub classifications is explored in this section. The efficiencies of the hubs were tested in three models to identify varying degrees of inefficiency. The first model considered is the BND_CCR model, given in model 1(a). This model shows no significant differences between any of the pairwise comparisons of the groups. The lone exception to this observation is the comparison of the small hub and medium hub that yields a *p*-value of 0.021, which is significant for many significance levels. An examination of the mean ordering reveals that the medium and large hubs have the best efficiency scores, which follows the prior results on RTS, which indicate that this group of hubs is more likely to comprise the constant RTS part of the frontier. The fact that there are no significant differences amongst the efficiency scores leads to the conclusion that the BND_CCR model does not have the ability to properly discriminate among the hub classifications. The resulting *p*-values from all three tests are given in Table 8.

For a more comprehensive result, the BND_BCC model is run. The major difference in this experiment is the inclusion of the convexity constraint

Table 10. Mean Ordering of Efficiency Scores by Hub Classification.

BND_CCR		BND_BCC		BND_SBM	
Hub classification	Average efficiency score	Hub classification	Average efficiency score	Hub classification	Average efficiency score
Medium hubs	1.1947335	Non-hubs	1.110912	Small hubs	1.940573281
Large hubs	1.203890595	Large hubs	1.120096667	Medium hubs	1.9456933
Non-hubs	1.2077125	Medium hubs	1.1828435	Large hubs	2.196298929
Small hubs	1.208442656	Small hubs	1.202168594	Non-hubs	2.615318

$\Sigma\lambda = 1$ to the model in Eq. (1a), thus allowing for efficiency of hubs that display increasing or decreasing RTS. This modification resulted in significant differences in all pairwise comparisons except for two comparisons. The comparisons between the non-hub and large hub groups and the non-hub and medium hub groups are the lone exceptions. This result, in addition to the mean ordering of the efficiency scores as noted in Table 10, demonstrates that the efficiency score of the small hubs is clearly the lowest among all the classes and that the non-hubs benefit the most from the convexification of the frontier. Whereas in the BND_CCR model, the non-hubs are ranked last in mean ordering, they are now ranked first and are statistically significantly better than the small hub group.

The final analysis uses a Bounded Slack-based Measurement (BND_SBM) model that measures efficiency based on the amount of increase in outputs needed reach the frontier. This quantity is measured by output slack $s^+{}_j$, which is then normalized by the original data elements y_{j0} and summed in the objective function. These changes yield the model given in model 1 (c). The inclusion of slack into the efficiency score is used to give a more accurate representation of the ‘‘total inefficiency’’ in a particular hub. Once again, we use the Wilcoxon Rank Sum test to identify significant differences among all pairwise combinations except for two comparisons: the comparison between the non-hub and small hub groups and the medium hub and large hub groups. The mean ordering (Table 10) shows that the non-hub group suffers the most from the inclusion of slack into the efficiency measure and is ranked last among all the classifications. And conversely, the small hubs benefit the most by going from last amongst the classifications to first. Yet, the p-values indicate that there is no significant difference between the non-hub and small hub classifications, thus resulting in a pairing of two groups by statistical significant difference, the non-hub and small hub in addition to the medium and large hubs.

Table 11. Summary of Malmquist Indices, Frontier Shift, and Catch-Up Effect of Yearly Hub Classification.

Years	Class	Malmquist Index					Frontier Shift					Catch-Up Effect				
		N	Mean	Standard deviation	Minimum	Maximum	*N*	Mean	Standard deviation	Minimum	Maximum	*N*	Mean	Standard deviation	Minimum	Maximum
2002–2003	NH	5	0.96	0.06	0.86	1.00	5	1.08	0.14	0.83	1.18	5	1.26	0.25	1.00	1.55
2002–2003	SH	16	0.99	0.13	0.78	1.17	16	0.97	0.12	0.75	1.17	16	1.42	0.22	1.00	2.00
2002–2003	MH	25	0.97	0.15	0.43	1.26	25	1.08	0.28	0.88	2.35	25	1.27	0.29	0.54	1.75
2002–2003	LH	21	1.02	0.12	0.78	1.44	21	1.01	0.09	0.85	1.30	21	1.25	0.22	1.00	1.85
2003–2004	NH	5	1.06	0.19	0.83	1.33	5	0.87	0.49	0.26	1.48	5	1.35	1.02	0.53	3.02
2003–2004	SH	16	1.29	0.43	0.72	2.31	16	0.86	0.27	0.45	1.39	16	1.15	0.38	0.67	1.95
2003–2004	MH	25	0.84	0.20	0.53	1.34	25	1.24	0.28	0.74	1.98	25	0.82	0.19	0.44	1.23
2003–2004	LH	21	0.88	0.26	0.48	1.59	21	1.22	0.36	0.64	2.07	21	0.86	0.25	0.40	1.38
2004–2005	NH	5	0.96	0.10	0.78	1.02	5	1.12	0.38	0.78	1.77	5	1.25	0.46	0.64	1.89
2004–2005	SH	16	1.04	0.28	0.60	1.70	16	1.00	0.24	0.59	1.53	16	1.15	0.30	0.66	1.59
2004–2005	MH	25	1.09	0.26	0.69	1.95	25	0.94	0.17	0.51	1.33	25	1.18	0.29	0.67	1.99
2004–2005	LH	21	1.12	0.34	0.76	2.49	21	0.95	0.18	0.40	1.33	21	1.07	0.30	0.75	2.22
20**02**–20**05**	**NH**	***5***	**0.80**	**0.24**	**0.42**	**1.00**	**5**	**1.22**	**1.07**	**0.26**	**2.94**	**5**	**2.36**	**2.19**	**0.38**	**5.88**
20**02**–20**05**	**SH**	**16**	**1.16**	**0.61**	**0.55**	**2.83**	**16**	**0.94**	**0.32**	**0.36**	**1.47**	**16**	**1.83**	**0.69**	**0.72**	**3.22**
20**02**–20**05**	**MH**	**25**	**0.81**	**0.25**	**0.38**	**1.22**	**25**	**1.22**	**0.43**	**0.69**	**2.77**	**25**	**1.20**	**0.39**	**0.42**	**1.75**
20**02**–20**05**	**LH**	**21**	**0.96**	**0.35**	**0.44**	**1.77**	**21**	**1.17**	**0.48**	**0.64**	**2.83**	**21**	**1.10**	**0.28**	**0.50**	**1.79**

Note: The bold text is an indication of non-consecutive years, particularly the 1st and last year of the dataset.

5.4. Malmquist Indices/Efficiency Change

In the years following the events of September 11, the airline industry faced major changes. We attempted to understand more about the affects of these changes between 2002 and 2005 using the Malmquist index. The Malmquist index is decomposed into two components: the FS and the CU effect. Each of these two components shows different aspects of the changes in efficiency. The FS gives an indication of how the overall industry has changed over time, while the CU shows the change in efficiency of the hub.

The time period for the comparison of the Malmquist index is completed on two different groupings. The first grouping compares the difference in performance in the year 2002 and the year 2005. This gives insight into how the airline industry has changed in total over the entire four-year time period. The second grouping is a year-by-year comparison examining the pairwise comparisons of 2002–2003, 2003–2004, and 2004–2005. This comparison helps to decide exactly where in the time window the change occurs during the selected time period. The summary of these results are listed in Table 11.

The result of the first comparison (2002–2005) shows no significant differences among the efficiency scores of the hub classifications except in the CU between the large and small hubs. The small hubs are statistically better than the large hubs from the years 2002 to 2005. This shows that the small hubs have done a better job at recovering in airport efficiency during this time period.

When the years are paired in the second grouping to determine exactly when the efficiency change occurs, the pair 2003–2004 shows results that indicate a significant change. The Wilcoxon Rank Sum test shows a statistically significant difference in the Malmquist Index, CU, and FS between the years 2003 and 2004. The small hubs are statistically different than both the medium and the large hubs, thus giving further proof that the small hubs did a better job in recovering from the September 11 effect.

6. CONCLUDING REMARKS AND FUTURE RESEARCH

In this chapter, we analyzed the performance of major airports in the United States using DEA. First, we found that significant differences among hub and non-hub airports do exist by using a non-radial-based DEA approach that decomposes the efficiency scores into scale efficiency, technical efficiency, and mixed efficiency. Second, this chapter examined the change in the efficiency of airports between the years of 2002 and 2005 and is able to show a significant improvement in both the efficient operations of the individual airports and also an increase in the efficiency of the entire industry. We emphasize that we include on-time operation in our model, which is a key factor in both customer satisfaction and efficient operations.

In future research, we plan to extend this approach to include multiple additional factors that affect on-time performance of airports (security delays, inclement weather, etc.). These factors have previously been identified as a critical factor that affects various types of inefficiencies in airport operations. Additionally, multiple perspectives of airport efficiency should be studied to understand the fundamentals that allow an airport to be attractive for an airline and neighboring or partnering businesses. And finally, to understand the cascading effect of delays in airports, a network-based approach will be needed to identify the origin sources of delays and methods to prevent catastrophic propagation throughout airline networks.

REFERENCES

Abdelghany, K. F., Shah, S. S., Raina, S., & Abdelghany, A. F. (2004). A model for projecting flight delays during irregular operation conditions. *Journal of Air Transportation Management*, *10*, 385–394.

Adler, N., & Berechman, J. (2001). Measuring airport quality from the airlines' viewpoint: An application of data envelopment analysis. *Transport Policy*, *8*, 171–281.

Adler, N., & Golany, B. (2001). Evaluation of deregulated airline networks using data envelopment analysis combined with principal component analysis with and application to Western Europe. *European Journal of Operational Research*, *132*, 260–273.

Balk, B. M. (2001). Scale efficiency and productivity change. *Journal of Productivity Analysis*, *15*, 159–183.

Bazargan, M., & Vasigh, B. (2003). Size versus efficiency: A case study of US commercial airports. *Journal of Air Transport Management*, *9*(3), 187–193.

Bethune, G. (2005). How to fix the airline mess. *Time*, 55, September 26.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, *2*, 429–444.

Cooper, W. W., Seiford, L., & Tone, K. (2006). *Data envelopment analysis* (2nd ed.). Springer.

Fernandes, E., & Pacheco, R. R. (2002). Efficient use of airport capacity. *Transportation Research Part A*, *36*, 225–238.

Gillen, D., & Lall, A. (1997). Developing measures of airport productivity and performance: An application of data envelopment analysis. *Transportation Research Part E*, *33*(4), 261–273.

Parker, D. (1999). The performance of BAA before and after privatization. *Journal of Transport Economics and Policy*, *33*(Pt. 2), 133–146.

Pels, E., Nijkamp, P., & Ritveld, R. (2003). Inefficiencies and economies of European airport operations. *Transportation Research Part E*, *39*, 341–361.

Ray, S. C., & Delsi, E. (1997). Productivity growth, technical progress, and efficiency change in industrialized countries: Comment. *The American Economic Review*, *87*, 1033–1039.

Sarkis, J. (2000). An analysis of the operational efficiency of major airports in the United States. *Journal of Operations Management*, *18*, 335–351.

Seiford, L., & Zhu, J. (1999). An investigation of returns to scale in data envelopment analysis. *Omega*, *27*, 1–11.

Tone, K. (2001). A slacks-based measure of efficiency in data envelopment analysis. *European Journal of Operational Research*, *130*(3), 489–509.

Tone, K. (2004). Malmquist productivity index. In: W. Cooper, L. Seiford & J. Zhu (Eds), *Handbook on data envelopment analysis* (ch. 8, pp. 203–227). Norwell, MA: Springer.

U.S. Department of Transportation. (2005). Research and Innovation Technology Administration (RITA). *Bureau of Transportation Statistics*. Available at http://www.bts.gov

USING DATA ENVELOPMENT ANALYSIS TO ANALYZE THE PERFORMANCE OF NORTH AMERICAN CLASS I FREIGHT RAILROADS

Rashmi Malhotra, D. K. Malhotra and Harvey Lermack

ABSTRACT

With increased crude oil prices, railroad is emerging as a cheaper alternative to trucks and other less fuel efficient modes of transportation. As a result, with increase in crude oil price, while other modes of transportation have suffered economic slump, railroad industry is thriving with every company reporting an increase in revenue and profits. In this study, we analyze the performance of seven North American Class I freight railroads. In this chapter, we illustrate the use of data envelopment analysis (DEA), an operations research technique, to analyze the financial performance of the U.S. railroad industry by benchmarking a set of financial ratios of a firm against its peers. DEA clearly brings out the firms that are operating more efficiently in comparison with other

Financial Modeling Applications and Data Envelopment Applications
Applications of Management Science, Volume 13, 113–131

ISSN: 0276-8976/doi:10.1108/S0276-8976(2009)0000013009

firms in the industry and points out the areas in which poorly performing firms need to improve.

1. INTRODUCTION

With increase in crude oil prices and the consequent increase in fuel costs for different modes of transportations, railroad industry is back in focus as a popular means of transportation for people as well as for goods. Freight trains, too, are moving into a new, more promising future. The nation needs an economical way to move its burgeoning volume of imports, and trains can do the job better than trucks. Trains use fuel more efficiently and avoid the costly delays caused by traffic. And of course they are also greener than smoke-belching 18-wheelers. This chapter addresses the financial performance of seven North American Class I freight railroads.

The North American freight railroad industry comprises over 550 railroads with 173,000 miles of track and earns about $54 billion in annual revenues.[1] Industry participants are segmented into classes, to which individual railroads are assigned based on revenues. This chapter analyzes the seven largest railroads – those in Class I, with over $346.8 million in 2006 revenues each – which account for 93% of the total industry revenues. In addition, there are 33 regional railroads and over 510 local (shortline or switching and terminal) railroads.[2]

During the 1960s and 1970s, the railroad industry experienced substantial financial strain, including the loss of interstate passenger rail service to airlines and interstate shipments to trucks, and bankruptcies of many of the largest carriers. During the 1970s, government financial participation was required to keep the industry afloat.

Before 1980, the highly unprofitable industry was regulated by the Interstate Commerce Commission (ICC), which approved routes, rates, and many operating activities. In 1980, the Staggers Act substantially deregulated the industry, permitting railroads to contractually establish rates and routes directly with shippers. This has led to two significant trends during the last three decades – resurgence of industry financial health and profitability, resulting in the end of all public financial participation in the industry; and significant consolidation of the mature industry. In fact, since 1981, the number of Class I railroads has dropped from 40 to 7.

All Class I railroads are privately owned, meaning they must compete in the open marketplace for financing. They are capital-intensive companies – in fact, railroads spend more than $20 billion per year (or 37% of revenues) to maintain and expand their track and equipment.[3] This places them at a disadvantage with air and truck transport modes, whose infrastructure is largely funded by the public. Therefore, it is important that railroads provide sufficient returns to investors to ensure a flow of enough capital at a reasonable cost of capital. However, while railroad profits have improved substantially since deregulation, the industry has struggled to provide returns that exceed their cost of capital after reinvestment in track and equipment.

Several recent trends and events have provided increased scrutiny of the industry and of its financial performance.

- Increased imports from Asia during the past 20 years have shifted and increased traffic flows of container traffic, especially east to west. In fact, Intermodal container freight recently became the largest category of freight moved by Class I's railroads.
- Recognizing the improved financial climate surrounding the industry, Warren Buffett announced during 2007 that his company, Berkshire Hathaway, had made substantial investments in three Class I railroads – Norfolk Southern, Union Pacific, and Burlington Northern Santa Fe.[4] In keeping with his high profile in the investment community, other investors immediately turned their attention to the improving fortunes of the industry.
- In keeping with a trend to greater shareholder involvement in corporate governance, the private Children's Investment Fund (TCI), a significant holder of CSX stock, during October 2007 "urged CSX, the railroad operator, to overhaul its management structure in an attempt to rein in spending and improve its financial performance."[5] To date, this has resulted in two TCI-sponsored members being added to the CSX Board, and significant pressure being placed on the company to increase its financial performance.
- Since rail is a very fuel-efficient mode of freight transportation, recent increases in crude oil prices have led to increased shipments and additional revenues for the industry.
- As a result of deregulation, some shippers now have access to only a single railroad for their shipments. As a result, there have been calls for re-regulation, especially to oversee service and rates to these "captive" shippers.

This chapter seeks to utilize data envelopment analysis (DEA) to investigate the performance of the Class I railroad operators. The resulting data will be of interest to railroad operators, investors, capital suppliers, and the academic community.

Rest of the chapter is organized along the following lines. Section 2 provides a review of previous studies on DEA applications in performance evaluation. Section 3 discusses the data and methodology used in this study. Section 4 provides an empirical analysis of our results. Section 5 summarizes and concludes our study.

2. LITERATURE REVIEW

A number of studies have been published on different aspects of financial statement analysis. We include only those studies that use DEA in either financial statement analysis or analysis of financial performance of firms. Zhu (2000) uses DEA to develop a multi-factor financial performance model that recognizes tradeoffs among various financial measures. Kao and Liu (2004) compute efficiency scores based on the data contained in the financial statements of Taiwanese banks. They use these data to make advanced predictions of the performance of 24 commercial banks in Taiwan. Pille and Paradi (2002) analyze the financial performance of Ontario credit unions. They develop models to detect weaknesses in Credit Unions in Ontario, Canada. Yasar and McCure (1996) use DEA for measuring and assessing the financial performance for hospitals. They compute a financial performance index (FPI) as a measure of aggregate financial performance. They show that FPI across many financial ratios eases the comparison of an individual hospital with its peers. Feroz, Kim, and Raab (2003) are the only study that directly talks about financial statement analysis using DEA methodology. They show that DEA can augment the traditional ratio analysis to a consistent and reliable measure of managerial or operational efficiency of a firm. Halkos and Salamouris (2004) explore the efficiency of Greek banks with the use of a number of suggested financial efficiency ratios for the time period 1997–1999. They show that DEA can be used as either an alternative or a complement to ratio analysis for the evaluation of an organization's performance.

With regard to the railroad industry, most of the studies have analyzed the productivity and efficiency of various aspects of the industry. Hoon and Chunyan (1994) analyzed the productive efficiency of the railway services in 19 Organization for Economic Cooperation and Development (OECD) countries. They report that railway systems with high dependence on public subsidies are less efficient than similar railways with less dependence on subsidies. Cowie and Riddington (1996) evaluate the efficiency of the European railways through the use of a production frontier approach. Yu and Lin (2008) uses a multi-activity network DEA model to simultaneously estimate passenger and freight technical efficiency, service effectiveness, and technical effectiveness for 20 selected railways for the year 2002.

In this chapter, we extend previous studies by illustrating the use of DEA models to benchmark the performance of North American class I freight railroads in terms of financial performance. No previous study has benchmarked railroad firms in terms of financial performance.

3. DATA AND METHODOLOGY

The DEA (Charnes, Cooper, & Rhodes, 1978) is a widely used optimization-based technique that measures the relative performance of decision-making units (DMUs) that are characterized by a multiple objectives and/or multiple inputs structure. DEA[6] is a technique used to assess the comparative efficiency of homogenous operating units such as schools, hospitals, utility companies, sales outlets, prisons, and military operations. More recently, it has been applied to banks (Haslem, Scheraga, & Bedingfield, 1999) and mutual funds (Haslem & Scheraga, 2003; Galagedera & Silvapulle, 2002; McMullen & Strong, 1998; Murthi, Choi, & Desai, 1997). It is a powerful technique for measuring performance because of its objectivity and ability to handle multiple inputs and outputs that can be measured in different units. The DEA approach does not require specification of any functional relationship between inputs and outputs, or a priori specification of weights of inputs and outputs. DEA provides gross efficiency scores based on the effect of controllable and uncontrollable factors.

The DEA methodology measures the performance efficiency of organization units called DMUs. This technique aims to measure how efficiently a DMU uses the resources available to generate a set of outputs. The performance of DMUs is assessed in DEA using the concept of efficiency or productivity defined as a ratio of total outputs to total inputs. Efficiencies estimated using DEA are relative, that is, relative to the best-performing DMU or DMUs (if multiple DMUs are the most efficient). The most efficient DMU is assigned an efficiency score of unity or 100%, and the performance of other DMUs vary between 0% and 100% relative to the best performance.

We used the financial data available from *Hoovers Online* for this study. We used 8 financial ratios to evaluate 7 North American class I freight railroads. Seven companies that we include in our study are: Burlington Northern Santa Fe, Canadian National Railway, Canadian Pacific Railways Limited, CSX, Kansas City Southern, Norfolk Southern, and Union Pacific. These are the seven largest railroads – those in Class I, with over $346.8 million in 2006 revenues each – which account for 93% of the total industry revenues. We benchmark the financial performance of these companies on the basis of the following financial variables:

- *Average Collection Period* – Number of days on an average it takes for the company to receive payments owed in terms of receivables from its customers and clients.
- *Cash flow per share* – Cash flow is calculated as net income − preferred dividends + depreciation. It is divided by shares outstanding from the most recent balance sheet.
- *Quick Ratio* – Quick ratio equals cash and equivalents plus receivables divided by total current liabilities from the most recent balance sheet. Quick ratio measures a company's short-term liquidity.
- *Inventory Turnover Ratio* – Inventory turnover equals the cost of goods sold divided by the average inventory from the most recent balance sheet and the corresponding balance sheet a year ago. Inventory turnover measures inventory management efficiency.
- *Long-Term Debt Per Share* – Long-term debt per share equals long-term debt divided by shares outstanding from the most recent balance sheet.
- *Return on equity* – Return on equity equals the net income from total operations divided by common stock equity from the most recent balance sheet. It measures the return on each dollar invested by the common shareholders in a company.
- *Return on assets* – Return on assets equals the net income from total operations divided by the total assets from the most recent balance sheet. A measure of profitability, ROA measures the amount earned on each dollar invested in assets.
- *Interest rate coverage ratio* – Interest coverage equals income before interest and taxes divided by the interest expense.

Table 1 illustrates the pooled data of the seven companies used for analysis.

Table 1. Pooled Data Set of North American Class I Freight Railroad Companies for Year 2008.

Company	Average Collection Period	Long-Term Debt per Share	Cash Flow per Share	Return on Equity	Return on Assets	Inventory Turnover	Quick Ratio	Interest Coverage Ratio
Burlington Northern Santa Fe (BNI)	24.96	24.35	10.51	0.168	0.054	12.5	0.5	6.63
Canadian National (CNI)	20.24	11.27	3.08	0.234	0.099	23.5	0.6	9.95
Canadian Pacific (CP)	47.41	29.02	8.84	0.219	0.083	11.4	0.7	6.33
CSX Corp (CSX)	49.54	18.14	6.02	0.165	0.058	28.3	1.2	5.67
Kansas City Southern (KSU)	88.44	16.29	4.97	0.113	0.039	6.4	0.6	2.97
Norfolk Southern (NSC)	45.41	16	6.01	0.154	0.058	22.3	0.8	6.24
Union Pacific (UNP)	25.91	0.1475	6.56	0.129	0.052	12.7	0.7	7.35

3.1. Data Envelopment Model Specifications for the Railroad Industry

Besides the mathematical and computational requirements of the DEA model, there are many other factors that affect the specifications of the DEA model. These factors relate to the choice of the DMUs for a given DEA application, selection of inputs and outputs, choice of a particular DEA model [e.g., constant returns to scale (CRS), variables return to scale (VRS)] for a given application, and choice of an appropriate sensitivity analysis procedure (Ramanathan, 2003). Owing to DEA's non-parametric nature, there is no clear specification search strategy. However, the results of the analysis depend on the inputs/outputs included in the DEA model. There are two main factors that influence the selection of DMUs – homogeneity and the number of DMUs. To successfully apply the DEA methodology, we should consider homogenous units that perform similar tasks and accomplish similar objectives. In our study, the companies are homogenous as they are identified by *Hoovers Online* to be competitors.

Furthermore, the number of DMUs is also an important consideration. In addition, the number of DMUs should be reasonable so as to capture high-performance units and sharply identify the relation between inputs and outputs. The selection of input and output variables is the most important aspect of performance analysis using DEA. In general, the inputs should reflect the level of resources used or a factor that should be minimized. The outputs reflect the level of the economic variable factor and the degree to which an economic variable contributes to the overall strength (efficiency) of a company. There are some simple rules of thumb that guide the selection of inputs and outputs, and the number of participating DMUs.[7]

To study the performance of the railroad industry, we consider eight factors to develop the DEA model: average collection period, long-term debt per share, cash flow per share, return on equity, return on assets, inventory turnover, quick ratio, and interest rate coverage.

Of these 8 factors, we specify average collection period and long-term debt per share as input, because for a given company the lower these variables are the better the performance of the company is. Similarly, higher cash flow per share, return on equity, return on assets, inventory turnover, quick ratio, and interest rate coverage imply a better-performing company. Thus, we consider these variables as output variables. Finally, the choice of the DEA model is also an important consideration. We should select the appropriate DEA model with options such as input maximizing or output minimizing, multiplier or envelopment, and constant or variable returns to scale. DEA applications that involve inflexible inputs or not fully under control inputs should use output-based formulations. On the contrary, an application with outputs that are an outcome of managerial goals, input-based DEA formulations are more appropriate. In addition, for an application that emphasizes inputs and outputs, we should use multiplier version. Similarly, for an application that considers relations among DMUs, envelopment models are more suitable. Furthermore, the characteristics of the application dictate the use of constant or variable returns to scale. If the performance of DMUs depends heavily on the scale of operation, CRS is more applicable, otherwise variable returns to scale is a more appropriate assumption.

In our study, the comparative evaluation among the companies is an important consideration. Therefore, we select the envelopment models for our analysis. In addition, the outputs are an outcome of managerial goals. Therefore, input-based formulation is recommended for our study. The objective of the analysis is to suggest a benchmark for the railroad firms. Furthermore, to investigate the affect of scale of operations, if any, among the seven companies, we consider both variable returns to scale and constant returns to scale DEA models. Also, the structure of the DEA model (in envelopment form) uses an equation and separate calculation for every input and output. Therefore, all the input and output variables can be used simultaneously and measured in their own units. In this study, we use the input-oriented VRS to evaluate the efficiency of seven companies for the second quarter of 2008.

4. EMPIRICAL ANALYSIS

Each of the railroad company is a homogenous unit, and we can apply the DEA methodology to assess the comparative performance of these companies. We analyze and compute the efficiency of these companies using the financial statements for the second quarter of 2008. Table 2 illustrates the efficiency scores for seven companies. Furthermore, we also study the peers (model companies) for inefficient companies.

Table 2 shows the relative performance of the railroad companies benchmarked against each other. Table 2 also shows that five of seven companies were ranked as efficient based on the financials for the second quarter of 2008, and two companies were inefficient companies. Burlington Northern Santa Fe, Canadian National, Canadian Pacific, CSX Corp, and Union Pacific are 100% efficient. On the contrary, Kansas City Southern and Norfolk Southern are inefficient. Fig. 1 shows the efficiency frontier graph of the pooled company data. The 100% efficient companies (blue

Table 2. DEA Efficiency Scores for the Railroad Companies.

Company	Efficiency (%)
Burlington Northern Santa Fe (BNI)	100
Canadian National (CNI)	100
Canadian Pacific (CP)	100
CSX Corp (CSX)	100
Kansas City Southern (KSU)	27
Norfolk Southern (NSC)	81
Union Pacific (UNP)	100

Note: A company with 100% score is considered the most efficient and a company with less than 100% score is considered inefficient. Efficiency scores is based on average collection period, long-term debt per share, cash flow per share, return on equity, return on assets, inventory turnover, quick ratio, and interest rate coverage.

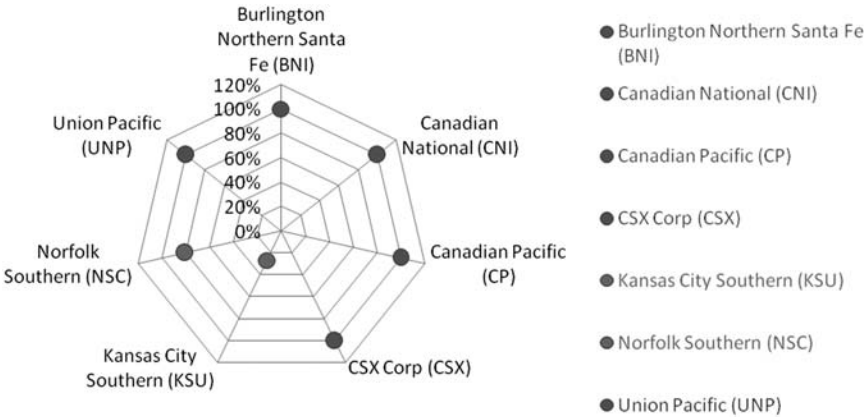


Fig. 1. Efficiency Frontier for the Benchmarked Companies.

dots) are on the efficiency frontier, whereas the inefficient companies (red dots) are inside the efficiency frontier. The DEA analyzer calculates the level of inefficiency by measuring the distance between the efficiency frontier and the inefficient companies. Therefore, a financial analyst can use this efficiency frontier to assess the relative efficiency of the firm in the industry. The DEA model compares the average collection period, long-term debt per share, cash flow per share, return on equity, return on assets, inventory turnover, quick ratio, and interest rate coverage.

We present the score in percentage value varying between 0% and 100%. We find that the input efficiency of Burlington Northern Santa Fe, Canadian National, Canadian Pacific, CSX Corp, and Union Pacific is 100%. On the contrary, the input efficiency of the remaining companies is Kansas City Southern (27%) and Norfolk Southern (81%). This means that the observed levels of cash flow per share, return on equity, return on assets, inventory turnover, quick ratio, and interest rate coverage for Kansas City Southern can be achieved with 27% of the current levels of average collection period and long-term debt per share. The same rationale applies to Norfolk Southern. Table 3 illustrates the efficiency scores and the corresponding ranking of the pooled companies in the year 2008. The average score is 87%, with five companies having efficiency levels above average while the remaining two are below the average level. Four 100% efficient companies turned out to be the best practices companies within the pooled database of the decision support system.

Table 3. Efficiency Score and Ranking of the Seven Companies for 2008.

Company	Efficiency (%)	Ranking
Union Pacific (UNP)	100	1
CSX Corp (CSX)	100	1
Canadian National (CNI)	100	1
Burlington Northern Santa Fe (BNI)	100	1
Canadian Pacific (CP)	100	1
Norfolk Southern (NSC)	81	2
Kansas City Southern (KSU)	27	3
Average	87	

Note: Ranking of individual company is based on the DEA efficiency scores from Table 3. Highest ranking is given to a company with the efficiency score of 100.

Table 4. Peer Companies and Their Weights in Percentage.

Company	Burlington Northern Santa Fe (BNI)	Canadian National (CNI)	CSX Corp (CSX)	Union Pacific (UNP)
Kansas City Southern (KSU)	0%	38%	0%	62%
Norfolk Southern (NSC)	8%	17%	50%	25%

Note: This table shows those companies that can serve as a benchmark for companies with DEA efficiency score of less than 100.

The best practices companies Burlington Northern Santa Fe, Canadian National, CSX Corp, and Union Pacific are 100% efficient. As Kansas City Southern and Norfolk Southern are inefficient, the next step is to identify the efficient peer group or companies whose operating practices can serve as a benchmark to improve the performance of these companies.

Table 4 illustrates the peer group for the inefficient companies.

As shown in Table 4, Canadian National and Union Pacific serve as peer for Kansas City Southern. In addition, Kansas City Southern is more comparable to Union Pacific (weight 62%) and less comparable to its more distant peer Canadian National (38%).Thus, Kansas City Southern should scale up its cash flow per share, return on equity, return on assets, inventory turnover, quick ratio, and interest rate coverage. Similarly, Norfolk Southern has CSX Corp (50%) as the closest peer that it should emulate and Union Pacific (25%) as the distant peer company that can also be investigated. Furthermore, Union Pacific has Canadian National (17%) as its far-distant peer and Burlington Northern Santa Fe (8%) as its furthest peer. Finally, Union Pacific and Canadian National are the most efficient company among the given pool of the companies in the DSS, as not only are Union Pacific and Canadian National 100% efficient, they also serve as the role model for all other companies. Similarly, CSX Corp is the next most efficient company among the group of companies. CSX Corp serves as the immediate peer for Norfolk Southern. The efficient peer companies have a similar mix of input–output levels to that of the corresponding inefficient company, but at more absolute levels. The efficient companies generally have higher output levels relative to the company in question. The features of efficient peer companies make them very useful as role models that inefficient companies can emulate to improve their performance. Furthermore, Union Pacific and Canadian National are the immediate efficient peers for the inefficient companies, so its frequency of use as an efficient-peer, expressed as a percentage of the number of pareto-inefficient companies, is 100%. Burlington Northern Santa Fe and CSX Corp serve as an immediate peer for one company. Thus, we have enhanced confidence that Burlington Northern Santa Fe and CSX Corp are genuinely well-performing companies as they outperform all the other companies. Furthermore, these companies are more likely to be a better role model for less efficient companies to emulate as their operating practices and environment match the majority of the other companies quite closely. Table 5 displays the benchmarking factor and the hit percentage of efficient company.

After calculating the efficiency of a company using DEA, and identifying the efficient peers, the next step in DEA is feasible expansion of the output or contraction of the input levels of the company within the possible set of input–output levels. The DEA efficiency measure tells us whether or not a given company can improve its performance relative to

Table 5. Benchmarking Factor and Hit Rate for Pareto Efficient Companies.

Company	Benchmarking Factor	Hit rate (%)
Burlington Northern Santa Fe (BNI)	1	100
Canadian National (CNI)	2	50
CSX Corp (CSX)	1	100
Union Pacific (UNP)	2	50

the set of companies to which it is being compared. Therefore, after minimizing the input efficiency, the next stage involves calculating the optimal set of slack values with an assurance that input efficiency will not decrease at the expense of slack values of the input and output factors. Once efficiency has been minimized, the model does seek the maximum sum of the input and output slacks. If any of the slack values is positive at the optimal solution to the DEA model, it implies that the corresponding output of the company (DMU) can improve further after its output levels have been raised by the efficiency factor, without the need for additional input. If the efficiency is 100% and the slack variables are zero, then the output levels of a company cannot be expanded jointly or individually without raising its input level. Furthermore, its input level cannot be lowered given its output levels. Thus, the companies are Pareto-efficient with technical output efficiency of 1. If the company is 100% efficient but one slack value is positive at the optimal solution, then the DEA model has identified a point on the efficiency frontier that offers the same level on one of the outputs as company A in question, but it offers in excess of the company A on the output corresponding to the positive slack. Thus, company A is not Pareto-efficient, but with radial efficiency of 1 as its output cannot be expanded jointly. Finally, if the company A is inefficient (<100%) or the efficiency factor is less than 1, then the company in question is not Pareto-efficient and efficiency factor is the maximum factor by which both its observed input levels can be reduced without changing its output. If at the optimal solution, we have not only input efficiency <1 but also some positive slack, then the output of company A corresponding to the positive slack can be raised by more than the factor output efficiency, without the need for additional input. The potential additional output at company A is not reflected in its efficiency measure because the additional output does not apply across all output dimensions. Table 6 illustrates the slack values identified in the next stage of the DEA. The slack variables for 100% efficient companies are zero. Therefore, Burlington Northern Santa Fe, Canadian National, Canadian Pacific, CSX Corp, and Union Pacific are Pareto-efficient as the DEA model has been unable to identify some feasible production point, which can improve on some other input or output level. On the contrary, for Kansas City Southern, there is further scope for increasing cash flow per share by .27, return on equity by .06 units, return on assets by .03 units, inventory turnover by 10.41 units, quick ratio by .06 units, and interest coverage by 5.37 units. Kansas City Southern can follow Canadian National and

Table 6. Slack Variables for Inefficient Companies (Efficiency < 100%) (2008).

Company	Cash Flow per Share	Return on Equity	Return on Assets	Inventory Turnover	Quick Ratio	Interest Coverage Ratio
Kansas City Southern (KSU)	0.27	0.06	0.03	10.41	0.06	5.37
Norfolk Southern (NSC)	0.00	0.01	0.01	0.00	0.12	0.66

Note: This table shows the adjustment needed in each of the seven economic variables for an inefficient company to become efficient.

Union Pacific as its role model and emulate their policies. Similarly, Norfolk Southern can increase its return on equity by .01 units, return on assets by .01 units, quick ratio by .12 units, and interest rate coverage by .66 units. Table 6 illustrates the slack values of the relevant factors for inefficient companies.

The next step in our analysis was to perform sensitivity analysis of the DEA model. DEA is an extreme point technique because the efficiency frontier is formed by actual performance of best-performing DMUs (Ramanathan, 2003). Furthermore, as DEA is a non-parametric technique, statistical hypothesis tests are difficult. It is possible for a DMU to obtain a value of utility by simply improving its performance in terms of only one particular output ignoring others. One way of checking the sensitivity of DEA efficiency of a DMU is by omitting one or more inputs or outputs. Thus, we used 26 different models to calculate efficiency of the railroad companies. Table 7 summarizes the results of our analysis. Table 8 displays the average efficiency, the standard deviation of the efficiencies, and median efficiency level for each country. Table 9 lists all the companies and their rankings based on average efficiency.

As expected, Canadian National and Union Pacific are the most efficient, followed closely by Union Pacific, Canadian Pacific, and CSX Corp. Norfolk Southern and Kansas City Southern are the most inefficient company.

5. SUMMARY AND CONCLUSIONS

With the increase in crude oil prices, railroad industry around the world is undergoing a major transformation. Although other modes of transportation

Table 7. DEA Models and Their Corresponding Efficiencies.

Company	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12
Burlington Northern Santa Fe (BNI)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Canadian National (CNI)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Canadian Pacific (CP)	1.00	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.53
CSX Corp (CSX)	1.00	0.45	0.45	0.45	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.48
Kansas City Southern (KSU)	1.00	0.24	0.24	0.24	0.24	0.24	0.24	1.00	1.00	1.00	1.00	0.27
Norfolk Southern (NSC)	0.72	0.49	0.49	0.49	0.69	0.69	0.72	0.72	0.72	0.81	0.81	0.52
Union Pacific (UNP)	1.00	0.87	0.87	0.87	0.87	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Company	Model 13	Model 14	Model 15	Model 16	Model 17	Model 18	Model 19	Model 20	Model 21	Model 22	Model 23	Model 24
Burlington Northern Santa Fe (BNI)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Canadian National (CNI)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Canadian Pacific (CP)	0.53	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.50	1.00
CSX Corp (CSX)	0.48	0.48	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.45	0.45
Kansas City Southern (KSU)	0.27	0.27	0.27	1.00	1.00	0.27	0.27	0.27	0.27	0.27	0.24	0.24
Norfolk Southern (NSC)	0.52	0.52	0.81	0.81	0.81	0.81	0.81	0.81	0.81	0.81	0.49	0.49
Union Pacific (UNP)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.87	0.87

Company	Model 25	Model 26	Average (%)
Burlington Northern Santa Fe (BNI)	1.00	1.00	100
Canadian National (CNI)	1.00	1.00	100
Canadian Pacific (CP)	1.00	1.00	93
CSX Corp (CSX)	0.45	1.00	81
Kansas City Southern (KSU)	0.24	0.24	46
Norfolk Southern (NSC)	0.49	0.69	67
Union Pacific (UNP)	0.87	0.87	96

Table 8. Average Percentage Efficiency of All Companies.

Company	Average (%)	Standard Deviation	Median (%)
Burlington Northern Santa Fe (BNI)	100	8.85E−12	100
Canadian National (CNI)	100	5.33E−11	100
Canadian Pacific (CP)	93	0.17784	100
CSX Corp (CSX)	81	0.263713	100
Kansas City Southern (KSU)	46	0.337206	27
Norfolk Southern (NSC)	67	0.136688	72
Union Pacific (UNP)	96	0.062841	100

Table 9. Company Rankings Based on Efficiency.

Company	Average (%)	Ranking
Canadian National (CNI)	100	1
Burlington Northern Santa Fe (BNI)	100	1
Union Pacific (UNP)	96	2
Canadian Pacific (CP)	93	3
CSX Corp (CSX)	81	4
Norfolk Southern (NSC)	67	5
Kansas City Southern (KSU)	46	6

have suffered economic slump due to higher crude oil prices, railroad industry is thriving with every company reporting an increase in revenue and profits. In this study, we analyze the performance of seven North American Class I freight railroads. In this chapter, we illustrate the use of DEA, an operations research technique, to analyze the financial performance of the U.S. railroad industry by benchmarking a set of financial ratios of a firm against its peers. DEA employs relative efficiency, a concept enabling comparison of companies with a pool of known efficient companies. The DEA model compares a firm with the pool of efficient companies by creating an *efficiency frontier* of good firms – a tolerance boundary created by establishing the efficiency of firms in terms of several sets of financial ratios. Companies lying beyond this boundary can improve one of the input values without worsening the others. We find that Burlington Northern Santa Fe, Canadian National, Canadian Pacific, CSX Corp, and Union Pacificare are 100% efficient. On the contrary, Norfolk Southern and Kansas City Southern are inefficient. We also illustrate the areas in which inefficient companies are lacking behind efficient firms.

We also provide an insight into the benefits of DEA methodology in analyzing financial statements of firms. The decision support system stores the company's historical data, competitive firm's data, and other industry-specific data, and uses the DEA methodology to analyze a firm's performance. Moreover, DEA modeling does not require prescription of the functional forms between inputs and outputs. DEA uses techniques such as mathematical programming that can handle a large number of variables and constraints. As DEA does not impose a limit on the number of input and output variables to be used in calculating the desired evaluation measures, it is easier for loan officers to deal with complex problems and other considerations they are likely to confront.

NOTES

1. Federal Railroad Administration, *Freight Railroading*, 2008, http://www.fra.dot.gov/us/content/4
2. *Ibid.*
3. Association of American Railroads, *Overview of America's Freight* Railroads, May 2008, http://www.aar.org/~/media/AAR/BackgroundPapers/775.ashx
4. Sorkin, Andrew Ross, "Buffett Discloses More Railroad Stakes," *New York Times*. 16 May 2007, http://dealbook.blogs.nytimes.com/2007/05/16/buffett-discloses-more-railroad-stakes/?scp = 2&sq = Buffett%20Union%20Pacific&st = cse
5. Activist Hedge Fund Presses CSX for Change, 17 October 2007 NYT, http://www.nytimes.com/2007/10/17/business/17hedge.html?_r = 1&scp = 7&sq = csx%20childrens%20fund&st = cse&oref = slogin.
6. For mathematical details of the data envelopment analysis model, see Zhu (2003).
7. The following are the guidelines for DMU model selection:

(a) The number of DMUs is expected to be larger than the product of number of inputs and outputs (Darrat, Can, & Yousef, 2002; Avikiran, 1984) to discriminate effectively between efficient and inefficient DMUs. The sample size should be at least 2 or 3 times larger than the sum of the number of inputs and outputs (Ramanathan, 2003).

(b) The criteria for selection of inputs and outputs are also quite subjective. A DEA study should start with an exhaustive, mutual list of inputs and outputs that are considered relevant for the study. Screening inputs and outputs can be quite quantitative (e.g., statistical) or qualitative that are simply judgmental, use expert advice, or use methods such as analytical hierarchy process (Saaty, 1980). Typically inputs are the resources utilized by the DMUs or condition affecting the performance of DMUs. On the contrary, outputs are the benefits generated as a result of the operation of the DMUs and record higher performance in terms of efficiency. Typically, we should restrict the total number of inputs and outputs to a reasonable level. As the number of inputs and outputs to a reasonable level. As the number of inputs and outputs increases, more number of DMUs get an efficiency rate of 1, as they become too specialized to be evaluated with respect to other units (Ramanathan, 2003).

REFERENCES

Avikiran, N. (1984). Investigating technical and scale efficiencies of Australian Universities through data envelopment analysis. *Socio-Economic Planning Sciences*, *35*, 57–80.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, *2*(6), 429.

Cowie, J., & Riddington, G. (1996). Measuring the efficiency of European railways. *Applied Economics*, *28*(8), 1027–1035.

Darrat, A., Can T., & Yousef, T. (2002). Assessing cost and technical efficiency of banks in Kuwait. Paper presented to the ERF's 8th Annual Conference in Cairo, ERF, Cairo, Egypt. Available at http://www.erf.org.eg/html/Finance_8th/Assessingcost-Darrat&Yousef.pdf

Feroz, E., Kim, S., & Raab, R. (2003). Financial statement analysis: A data envelopment analysis approach. *The Journal of the Operational Research Society*, *54*(1), 48–58.

Galagedera, D., & Silvapulle, P. (2002). Australian mutual fund performance appraisal using Data envelopment analysis. *Managerial Finance*, *28*(9), 60.

Halkos, G., & Salamouris, D. (2004). Efficiency measurement of the Greek commercial banks with the use of financial ratios: A data envelope analysis approach. *Management Accounting Research*, *15*(2), 201.

Haslem, J. A., & Scheraga, C. A. (2003). Data envelopment analysis of morningstar's large-cap mutual funds. *Journal of Investing*, *12*(4), 41.

Haslem, J. A., Scheraga, C. A., & Bedingfield, J. P. (1999). DEA efficiency profiles of U.S. banks operating internationally. *International Review of Economics & Finance*, *8*(2), 165.

Hoon, O., & Chunyan, Y. (1994). Economic efficiency of railways and implications for public policy. *Journal of Transport Economics and Policy*, *28*(2), 121–139.

Kao, C., & Liu, S. (2004). Predicting bank performance with financial forecasts: A case of Taiwan commercial banks. *Journal of Banking & Finance*, *28*(10), 2353.

McMullen, P. R., & Strong, R. A. (1998). Selection of mutual funds using data envelopment analysis. *The Journal of Business and Economic Studies*, *4*(1), 1.

Murthi, B. P. S., Choi, Y. K., & Desai, P. (1997). Efficiency of mutual funds and portfolio performance measurement: A non-parametric approach. *European Journal of Operational Research*, *98*(2), 408.

Pille, P., & Paradi, J. (2002). Financial performance analysis of Ontario (Canada) credit unions: An application of DEA in the regulatory environment. *European Journal of Operational Research*, *139*(2), 339–350.

Ramanathan, R. (2003). *An introduction to data envelopment analysis – a tool for performance measurement*. New Delhi, India: Sage Publications.

Saaty, T. L. (1980). *The analytical hierarchy process*. New York: McGraw-Hill.

Yasar, O., & McCure, M. (1996). Development of a financial performance index for hospitals: DEA approach. *The Journal of the Operational Research Society*, *47*(1), 18–27.

Yu, M., & Lin, E. (2008). Efficiency and effectiveness in railway performance using a multi-activity network DEA model. *Omega*, *36*(6), 1005–1017.

Zhu, J. (2000). Multi-factor performance measure model with an application to Fortune 500 companies. *European Journal of Operational Research*, *123*(1), 105–124.

Zhu, J. (2003). *Quantitative models for performance evaluation and benchmarking*. Kluwer's International Series, p. 13.

USING REGRESSION AND DATA ENVELOPMENT ANALYSIS (DEA) TO FORECAST BANK PERFORMANCE OVER TIME

Ronald K. Klimberg, Kenneth D. Lawrence,
Ira Yermish, Tanya Lal and Daniel Mrazik

ABSTRACT

Forecasting is an important tool used to plan and evaluate business operations. Regression analysis is one of the most commonly used forecasting techniques for this purpose. Often forecasts are produced based on a set of comparable units such as individuals, groups, departments, or companies that perform similar activities. We apply a methodology that includes a new independent variable, the comparable unit's data envelopment analysis (DEA) relative efficiency, into the regression analysis. In this chapter, we apply this methodology to compare the performance of commercial banks over a 10-year time period.

Financial Modeling Applications and Data Envelopment Applications
Applications of Management Science, Volume 13, 133–142

ISSN: 0276-8976/doi:10.1108/S0276-8976(2009)0000013010

INTRODUCTION

Quantitative forecasting techniques use historical data to predict future outcomes. Most quantitative forecasting techniques can be categorized into either time series or causal models. Time series forecasting techniques use only the time series data itself to build the models. These time series approaches isolate and measure the impact of trend, seasonal, and cyclical time series components. Causal models use a set of independent (predictor) variables, possibly including the time series components, believed to influence the forecasted dependent variable. One of the most popular causal model approaches is regression analysis. Regression techniques employ the least squares method to establish a statistical relationship between the dependent (forecasted) variable and the set of independent (predictor) variables.

In many forecasting situations, analysts must produce forecasts for comparable units. Comparable units could be individuals, groups of individuals, departments or business, and operational entities. In this analytical environment each comparable unit should be performing a similar set of tasks. For example, preparing forecasts for a number of corporate divisions will predict the sales results in future periods for similar products based on prior investments in similar marketing promotions. When applying regression analysis, the established statistical relationship is an average relationship using one set of weights assigned to the independent variables. However, when regression is applied to a set of comparable units, the relative weight of each of the independent variables will vary from comparable unit to comparable unit. For example, if advertising is an independent variable, one comparable unit might emphasize advertising more (or less) than other comparable units. As a result, the regression model could provide forecast estimates that are too high or too low.

In this chapter, we apply and extend some of our recent work in which we introduced a methodology that incorporates into the regression forecasting analysis a new variable that captures the unique weighting of each comparable unit (Klimberg, Lawrence, & Lawrence, 2008a; Klimberg, Lawrence, & Lawrence, 2005; Klimberg, Lawrence, & Lal, 2008b). This new variable is the relative efficiency of each comparable unit. It is generated by a non-parametric technique called data envelopment analysis (DEA). The DEA efficiency variable is a nonlinear variable that takes into account a set of weighted inputs and outputs. In each of our previous studies, the inclusion of this multivariate variable has improved the regression forecasting model.

The main objective of this chapter is to present the results of a longitudinal study applying this methodology.

In the next section, we provide a brief introduction to DEA. Subsequently, we discuss the methodology and present the results of applying our methodology to a data set of commercial banks. Finally, the conclusions and future extensions are discussed.

DATA ENVELOPMENT ANALYSIS

One of the major concerns of managers in evaluating the performance of an operation within any type of organization is efficiency. Efficiency measures whether resources are being put to good use. One dimension of the efficiency of an operation of any organization is the manner by which that organization selects and uses resources to produce its products. The more products produced for a given amount of resources the more efficient (i.e., less wasteful) is the operation. To evaluate the relative efficiency of comparable components, Charnes, Cooper, and Rhodes (1978) proposed an innovative quantitative technique that they named DEA.

DEA utilizes linear programming to produce measures of the relative efficiency of comparable decision-making units (DMUs) that employ multiple inputs and outputs. The DMU is the component of the organization being evaluated. For example, a hospital may use the technique to evaluate different care-giving units. DEA takes into account multiple inputs and outputs to produce a single aggregate measure of relative efficiency for each DMU. It requires only that the selected inputs and outputs be quantifiable. The technique can analyze these multiple inputs and outputs in their natural physical units without reducing or transforming them into some common unit of measurement. Finally, DEA evaluates all the DMUs and all their inputs and outputs simultaneously, conservatively identifying the sets of relatively efficient and inefficient DMUs. Thus, the solution of a DEA model provides a manager a summary with comparable DMUs grouped together and ranked by relative efficiency. Since the appearance of the seminal paper in 1978, there have been thousands of theoretical contributions and practical applications in various fields using DEA. DEA has been applied to many diverse areas such as health care, military operations, criminal courts, university departments, banks, electric utilities mining operations, and manufacturing productivity (Klimberg, 1998; Klimberg & Kern, 1992; Seiford, 1996; Seiford & Thrall, 1990).

For DEA, efficiency is defined as the ratio of weighted outputs to weighted inputs:

$$\text{Efficiency} = \frac{\text{weighted sum of outputs}}{\text{weighted sum of inputs}}$$

The DEA approach identifies the set of weights (all weights must be positive) that maximizes individually each DMU's efficiency while requiring the corresponding weighted ratios of the other DMUs to be less than or equal to one. A DMU is considered relatively inefficient if its efficiency rating is less than 1 ($E<1$). The degree of inefficiency for a DMU is measured relative to a set of more efficient DMUs. However, a DMU identified as being efficient ($E=1$) does not necessarily imply absolute efficiency. It is only relatively efficient as compared to the other DMUs that are being considered.

The Charnes, Cooper, and Rhodes (CCR) DEA model (Charnes et al., 1978) is a linear program that compares the ratio of weighted outputs to weighed inputs, that is, efficiency, for each comparable unit. The efficiency E_k of comparable unit k is obtained by solving the following linear formulation:

$$\begin{aligned} \max\ E_k &= \sum_{r=1}^{t} u_r Y_{rk} \\ \text{s.t.}& \\ &\sum_{i=1}^{m} v_i X_{ik} = 1 \\ &\sum_{r=1}^{t} u_r Y_{rj} - \sum_{i=0}^{m} v_i X_{ij} \leq 0 \quad j = 1, \ldots, n \\ &u_r, v_i \geq \varepsilon \quad \forall r, i \end{aligned} \tag{1}$$

where

Parameters
Y_{rj} = amount of output r for comparable unit j
X_{ij} = amount of input i for comparable unit j
t = the number of outputs
m = the number of inputs
n = the number of comparable units
ε = an infinitesimal value

Decision variables
u_r = the weight assigned to output r
v_i = the weight assigned to input i

REGRESSION FORECASTING METHODOLOGY

We developed a new regression forecasting methodology designed to be applied to a historical data set of multiple inputs and outputs variables from a set of comparable units (Klimberg et al., 2005). We enhance this methodology by improving the variable selection/reduction process and generalizing the regression analysis (Klimberg et al., 2008a, 2008b). The modified three-step regression forecasting methodology process is described later.

Dependent Variable Selection and Variable Reduction

Given a data set of comparable units, or in DEA terminology DMUs, the initial universe of input and output factors to be considered is likely to be rather large. First, one of the output variables must be designated as the principal (critical) variable to be forecasted, for example, sales, production, or demand. Having identified the variable to be forecasted, depending on the size of the data set, the number of input and output variables may have to be reduced. In general, the combined total of inputs and outputs included in the model should be no more than half the number of DMUs being compared in the analysis (Boussofiane, Dyson, & Thanassoulis, 1991; Golany & Roll, 1989). However, the usefulness of the efficiency values produced by solving a DEA model is only as good as the recognized importance and merit of the factors included in the model and used to produce them. The elimination of influential factors may significantly reduce the quality of the DEA results. Furthermore, some factors may be correlated with other factors, yet should be included because of their importance and what they measure. Unlike regression analysis, DEA does not have a "stepwise" option. Nevertheless, both techniques share the goal of the principle of parsimony, that is, building a model that includes the least number of variables which sufficiently explains some dependent variable.

To select the smallest number of the most important input and output factors from the total universe of possible variables, Golany and Roll (1989) suggest a three-step approach. The first step consists of obtaining expert judgments from decision makers as to the level of significance of the different input and output factors. Possible procedures that may be used to obtain these opinions include the Delphi method, expert panels, and focus groups. The second step is to apply traditional quantitative techniques such as regression and correlation analysis to the variables under consideration for inclusion in the DEA model. These techniques would identify strong or weak relationships between and among input and output variables. Variables demonstrating weak relationships would be eliminated from the analysis or aggregated with another variable. The final step is to actually perform some DEA analysis. Golany and Roll suggest using DEA with the input and output factors that remain after completing the first two steps. The weights assigned to each of the factors for each DMU would be observed. Those factors consistently receiving small values are considered for elimination from the model.

DEA Analysis

Using the significant variables from Step 1, a DEA model is created and executed for each comparable unit/DMU. The DEA model will produce an efficiency score that measures the relative efficiency of each of the comparable units/DMUs. In addition to these efficiency scores, the DEA model generates for each comparable unit a unique set of weights assigned to the set of inputs and outputs values. The DEA process attempts to find the set of weights that will maximize a comparable unit's efficiency. Therefore, the DEA model selects the best possible set of weights for each comparable unit. The variation of these weights from comparable unit to comparable unit allows each comparable unit to have their own unique freedom to emphasize the importance of each of their input and output variables. The efficiency score measures how well they do this. In the next step, we use these efficiency scores as surrogate measures of the unique emphasis of the variables and of the units' performance.

Regression Analysis

A stepwise regression is first executed using all the significant input variables and includes the DEA efficiency score. Depending on the stepwise regression model results, additional statistical and regression analysis may be warranted.

Since the data set we studied in this chapter has a relatively small number of inputs and outputs, we adjust our procedure and eliminate the initial stepwise regression. As a result, the first step is to run a DEA for each comparable unit. We use the resultant efficiency scores as surrogate measures of performance. Using a principal output variable as the regression-dependent variable, all the input variables plus the DEA efficiency score as regression-independent variables, we run the regression. This regression model with the DEA efficiency variable should be superior to the model without the DEA efficiency variable. Superiority would be demonstrated with a significantly lower standard error of the mean and increased R^2.

EXAMPLE

Seiford and Zhu (1999) applied DEA to the 55 U.S. commercial banks that appeared in the *Fortune* 1000 list in April 1996. The DEA input variables were the number of employees, assets, and stockholder's equity. The DEA output variables were revenue and profit. The selection of these variables was "based on *Fortune's* original choice of factors for performance characterization" (Seiford & Zhu, 1999). We retrieved the same *Fortune* 1000 list of U.S. commercial banks from 1995 to 2005. Thirteen banks were common over this time period. Ten of these thirteen banks are regional, and during this time period, almost all were mid-cap, profitable, and had a higher return on assets than the industry average.

Using the data for these thirteen banks, we ran 10 DEA models for 1995–2004. Table 1 lists some descriptive statistics of the DEA efficiency scores for these years. As shown in Table 1 for the 10-year period examined, at least half of the banks consistently had efficiency scores greater than 90%. The minimum scores were generally around 80%, except in 2004 where one bank dropped to 62.1%.

Using the DEA efficiency scores as an input and revenue as our primary output variable, we ran regression models for 1996–2005. The basic regression equation used was:

$$\begin{aligned}\text{Revenue}(t) = \text{Employees}(t-1) + \text{Assets}(t-1)\\ + \text{Equity}(t-1) + \text{DEA}(t-1)\end{aligned}$$

Table 1. Descriptive Statistics of the DEA Efficiency Scores for Each Year.

	Year									
	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
Minimum	85.89	73.13	78.89	77.69	80.40	81.75	86.72	84.47	81.41	62.10
Maximum	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
# 100	5	3	5	5	4	7	5	5	3	3
Average	96.78	91.44	93.30	92.81	93.00	94.72	96.12	96.07	91.97	90.57
Median	97.29	90.98	96.80	95.53	95.52	100.00	98.14	98.39	91.22	92.48

Table 2. Regression R^2 Values for Each Year for Both Models.

R^2	Year									
	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
NoDEA	97.58	97.57	96.42	91.41	95.59	93.82	98.53	95.70	97.29	94.99
w/DEA	99.09	99.57	96.76	93.89	98.88	95.20	99.55	96.49	97.47	98.50
Difference	1.51	2.00	0.34	2.48	3.29	1.38	1.02	0.79	0.18	3.51

Table 3. Regression Standard Errors for Each Year and for Both Models.

Standard Error	Year									
	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
NoDEA	333.47	343.29	478.24	669.27	517.96	575.61	236.38	402.67	349.88	645.26
w/DEA	216.69	153.24	482.42	598.84	276.53	538.25	139.03	385.65	358.57	374.16
Difference	−116.78	−190.05	4.18	−70.43	−241.43	−37.36	−97.35	−17.02	8.69	−271.1

where $t = 1996, \ldots, 2005$. We refer to this model using the DEA variable as *w*/DEA, and conversely, the same regression without the DEA efficiency score variable is referred to as NoDEA.

Tables 2 and 3 summarize the regression models results with R^2 values and standard errors. The R^2 values in all the models are extremely high with the NoDEA models averaging 95.89 and only have a minimum of 91.41. The R^2 values for the *w*/DEA models were consistently greater than the NoDEA models. The average improvement was only 1.65%, having already started with high R^2 values. In Table 3 the standard error values for the *w*/DEA models show significantly more improvement in eight of the ten years, averaging a 23.8% decrease in the standard errors.

Since the models with all three variables and with the DEA models, four variables, had such high R^2 values, we decided to re-examine the models with less variables. To identify which variables to include, we ran stepwise regressions for each year (without the DEA variable). Each variable was significant in five of the ten years. The combination of stockholder's equity and assets had at least one variable significant in each of the 10 years. Consequently, we re-ran the models with and without the DEA variable and including only stockholder's equity and assets. Tables 4 and 5 summarize the regression models results with R^2 values and standard errors.

Table 4. Regression R^2 Values for Each Year and for the Revised Models.

R^2	Year									
	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
NoDEA	97.2	96.4	93.9	90.3	94.6	91.8	89	87.8	85.7	86.2
DEA	98.8	98.5	94.3	92.9	97.6	93.4	90.6	88.6	85.7	91.3
Difference	1.6	2.1	0.4	2.6	3	1.6	1.6	0.8	0	5.1

Table 5. Regression Standard Errors for Each Year and for the Revised Models.

Standard Error	Year									
	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
NoDEA	339.71	396.14	592.51	675.12	541.75	628.32	612.82	642.13	762.85	1016.88
DEA	237.78	271.96	605.42	610.26	380.33	596.53	598.72	655.56	803.28	848.36
Difference	−101.93	−124.17	12.92	−64.85	−161.42	−31.78	−14.10	13.43	40.43	−168.52

The R^2 values remain high, now, averaging at about 91.29. The R^2 values for the *w*/DEA models were again consistently greater than the NoDEA models. The average improvement was slightly higher than the initial models though only 2.0%. The standard error values for the *w*/DEA models (Table 5) showed significantly more improvement in seven of the ten years, averaging a smaller (compare to the initial models) decrease of 11.5%.

CONCLUSIONS

In this chapter, we applied a new regression forecasting methodology to forecasting comparable units. This approach included in the regression analysis a surrogate measure of the unique weighting of the variables and of performance. This new variable is the relative efficiency of each comparable unit that is generated by DEA. The results of applying this new regression forecasting methodology including a DEA efficiency variable to a data set demonstrated that this approach does provide an enhanced approach to forecasting comparable units. We plan to perform further testing with other data sets from other industries, some with more comparable units, and possibly with lower R^2 values.

REFERENCES

Boussofiane, A., Dyson, R. G., & Thanassoulis, E. (1991). Applied data envelopment analysis. *European Journal of Operational Research*, *52*(1), 1–15.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring efficiency of decision making units. *European Journal of Operational Research*, *2*, 429–444.

Golany, B., & Roll, Y. (1989). An application procedure for DEA. *Omega*, *17*(3), 237–250.

Klimberg, R. K. (1998). Model-based health decision support systems: Data envelopment analysis (DEA) models for health systems performance evaluation and benchmarking. In: J. Tan (Ed.), *Health decision support systems* (pp. 99–126). New York, NY: Aspen.

Klimberg, R. K., & Kern, D. (1992). *Understanding data envelopment analysis (DEA)*. Working Paper, pp. 92–44. Boston University School of Management.

Klimberg, R. K., Lawrence, K. D., & Lawrence, S. M. (2005). Forecasting sales of comparable units with data envelopment analysis (DEA). In: *Advances in business and management forecasting* (Vol. 4, pp. 201–214). Amsterdam, The Netherlands: JAI Press.

Klimberg, R. K., Lawrence, K. D., & Lawrence, S. M. (2008a). Improved performance evaluation of comparable units with data envelopment analysis (DEA). In: *Advances in business and management forecasting* (Vol. 5, pp. 65–75). Amsterdam, The Netherlands: Elsevier Ltd.

Klimberg, R. K., Lawrence, K. D., & Lal, T. (2008b). *Using data envelopment analysis (DEA) to forecast bank performance*. NEDSI, March, New York.

Seiford, L. M. (1996). Data envelopment analysis: The evaluation of the state of the art (1978–1995). *The Journal of Productivity Analysis*, *9*, 99–137.

Seiford, L. M., & Thrall, R. M. (1990). Recent developments in DEA: The mathematical programming approach to frontier analysis. *Journal of Econometric*, *46*, 7–38.

Seiford, L. M., & Zhu, J. (1999). Profitability and marketability of the top 55 U.S. commercial banks. *Management Science*, *45*(9), 1270–1288.

CREATING AN INDEX OF VULNERABILITY TO SEVERE COASTAL STORMS ALONG THE NORTH SHORE OF BOSTON☆

Samuel J. Ratick, Holly Morehouse and Ronald K. Klimberg

ABSTRACT

A great deal of uncertainty accompanies predictions of the potential effects of global climate change on the coastal hazards associated with severe storms. One way to obviate the effects of this uncertainty on the design of policies is to understand the manner in which populations are currently vulnerable to these types of hazards. In this chapter, we develop a method for constructing a relative composite measure of vulnerability using data envelopment analysis (DEA). Through the application of this index, and one constructed using a weighted average, to

☆This chapter represents one research track within a larger collaborative project at Clark University. The authors are solely responsible, however, for any mistakes or errors in the use of the products of their collaboration and advice. The results presented in this chapter, however, are the sole responsibility of the authors and may not reflect the position of the Administration.

Financial Modeling Applications and Data Envelopment Applications
Applications of Management Science, Volume 13, 143–178

ISSN: 0276-8976/doi:10.1108/S0276-8976(2009)0000013011

four costal towns along Boston's North Shore, we demonstrate their potential usefulness to policy formulation and implementation. The DEA composite index is shown to complement the information provided by the weighted average and helps overcome some of its shortcomings such as assigning importance weights and masking of the influence of one or a subset of vulnerability attributes. Acknowledging the spatial implications of floodplain protection and mitigation efforts, the indices are constructed and analyzed at a number of different geographic scales.

INTRODUCTION

The current focus on the potential effects of global warming includes concerns about the consequences of coastal hazards associated with accelerated sea level rise and changing storm climates. Prediction of these consequences, even on relatively large geographic scales, is problematic due to the scientific uncertainty that pertains to the timing, extent, magnitude, and geographic pattern of possible changes (Ratick et al., 1992). Concomitantly, the formulation, acceptance, and implementation of policies designed to anticipate and address those consequences has proven to be quite difficult. These problems are exacerbated when one considers the implications of the different geographic scales at which hazard assessments and mitigative policies may be undertaken. One approach that may obviate some of the effects of high levels of uncertainty is to understand the manner in which areas are currently vulnerable to coastal hazards and to formulate policies that attempt specifically to address those vulnerabilities. In this chapter, we develop a method for constructing indices to measure and map areas whose populations may be vulnerable to coastal hazards associated with severe storms. We construct (using census data at various geographic scales), map, and analyze the information provided by that index for four towns along the Massachusetts coast north of Boston.

Although vulnerability is an intuitively simple notion, it is surprisingly difficult to define and even more difficult to operationalize. It is described in the literature in numerous and sometimes inconsistent ways. Definitions of vulnerability range from a focus on physical exposure (Mitchell, 1989; Schneider & Chen, 1980; Barth & Titus, 1984; Solow & Ratick, 1994; Manson, Ratick, & Solow, 2002), to measures of socioeconomic status and access to resources (Susman, O'Keefe, & Wisner, 1983; Timmerman, 1981;

Cannon, 1994), to sociological investigations of the differential ability of groups to resist harm and to recover afterwards (Drabek, 1986; Bolin, 1982; Quarentelli, 1991), and to discussions of how the "hazard of place" concatenates with social profiles (Dow, 1992; Cutter, 1996; Cutter, Mitchell, & Scott, 2000). While research is currently on-going on this definitional scheme, defining vulnerability as the exposure of a population to a hazard together with the effected populations sensitivity to, and ability to cope with or adapt to, the hazard provide a workable general conceptualization (Dow, 1992; Clark et al., 1998; Brooks Nick, Adger, & Kelly, 2005). The difficult task facing public decision makers is to translate this vague and ambiguous concept into concrete and definite policy decisions knowing that life and property depend on the effectiveness of those decisions.

Two general concerns guide our development of the indices. One is that it is difficult to make decisions when dealing with numerous and disparate dimensions. Policy formulation may require a composite index that effectively aggregates the various dimensions of vulnerability to provide an assessment of the manner in which different areas or populations may be compared, and to address the question of who is vulnerable. Another is that reducing multiple dimensions, each providing a slightly different understanding of the issue, into a single measure may mean the loss or masking of potentially valuable information. In this context, a policy maker must be able to look at the index results and understand why or how a certain area may be more vulnerable to the effects of coastal storms. That is, an index should maintain some of the richness of the data it is summarizing.

In this chapter, we evaluate two index construction methods: one using the standard linear weighted average as an absolute measure of vulnerability and the other using data envelopment analysis (DEA) to construct a relative index measure. We demonstrate how the DEA index provides a scalar measure that allows for comparisons between places and that preserves information contained within the different dimensions of vulnerability. The first section of the chapter describes the geographic context of the case study area, the four towns on the north shore of Boston. The next section describes the choice of vulnerability attributes and their associated variable measures and maps each of the variables for the four towns. The development of the composite index using DEA is presented in the following section and includes an illustrative example. In the next section, the composite indices of vulnerability are constructed, mapped, and analyzed for the case study area at various geographic scales. The last section provides a summary and conclusions.

BACKGROUND OF THE CASE STUDY AREA

The case study area is composed of four coastal towns on the North Shore of Massachusetts: Saugus, Lynn, Revere, and Malden. Approximately five to ten miles north of Boston, portions of the four towns lie along the coast and fall within the Saugus-Pines River Basin (Fig. 1). The Saugus-Pines River Basin was chosen as the case study area because it offers unique research opportunities with a rich history of data collection and a strong scientific base in the area. Following the "Blizzard of 1978," the U.S. Army Corps of Engineers complied an eight-volume flood hazard reduction study specific to the four towns of Lynn, Malden, Revere, and Saugus (U.S. Army Corps of Engineers, 1990). The 4,000 acres of residential, industrial, and commercially developed land and tidal wetlands in the study area suffer frequently from coastal flooding. The area is crisscrossed with major transportation arteries, including major highways, railroads, commuter rails, and public transit lines, as well as utilities that serve Boston's North Shore. In the past thirty years, many storms have hit the area including major floods said to be in 10- to 100-year recurrence frequency, and smaller storms disrupt the area yearly. The worst storm on record, the "Blizzard of 1978," caused record flood depths of up to seven feet, forcing the evacuation of over 4,000 people and affecting the whole residential, commercial, industrial, and commuter population on the North Shore. In the period from 1990 to 1993, four storms with magnitudes that some have estimated to be equal to a "100 year storm" have hit the area: Hurricane Bob and the Halloween Northeaster in 1991, the Blizzard of December 1992, and the Blizzard of March 1993.

Without some intervention, these damages are forecast to occur in the future. The U.S. Army Corps of Engineers estimates damages at over \$100 million if a storm tide equal to that of the 1978 storm were to reoccur. Furthermore, they predicted at the time of their study that "(t)he worst coastal storm reasonably likely to hit the area, the Standard Project Northeaster (SPN), could cripple the region, causing upwards of 10 feet of flooding and \$500 million in damages-closing a General Electric plant, an important defense facility; affecting up to 5,000 residential, commercial, industrial, and public buildings; threatening utilities serving the North Shore and disrupting the lives of over 300,000 residents and employees in these communities and commuters who use the major transportation arteries which traverse this urban floodplain" (U.S. Army Corps of Engineers, 1990). These consequences do not include consideration of accelerated sea level rise or changing storm climates that potentially could occur in response to global warming (Ratick et al., 1992).

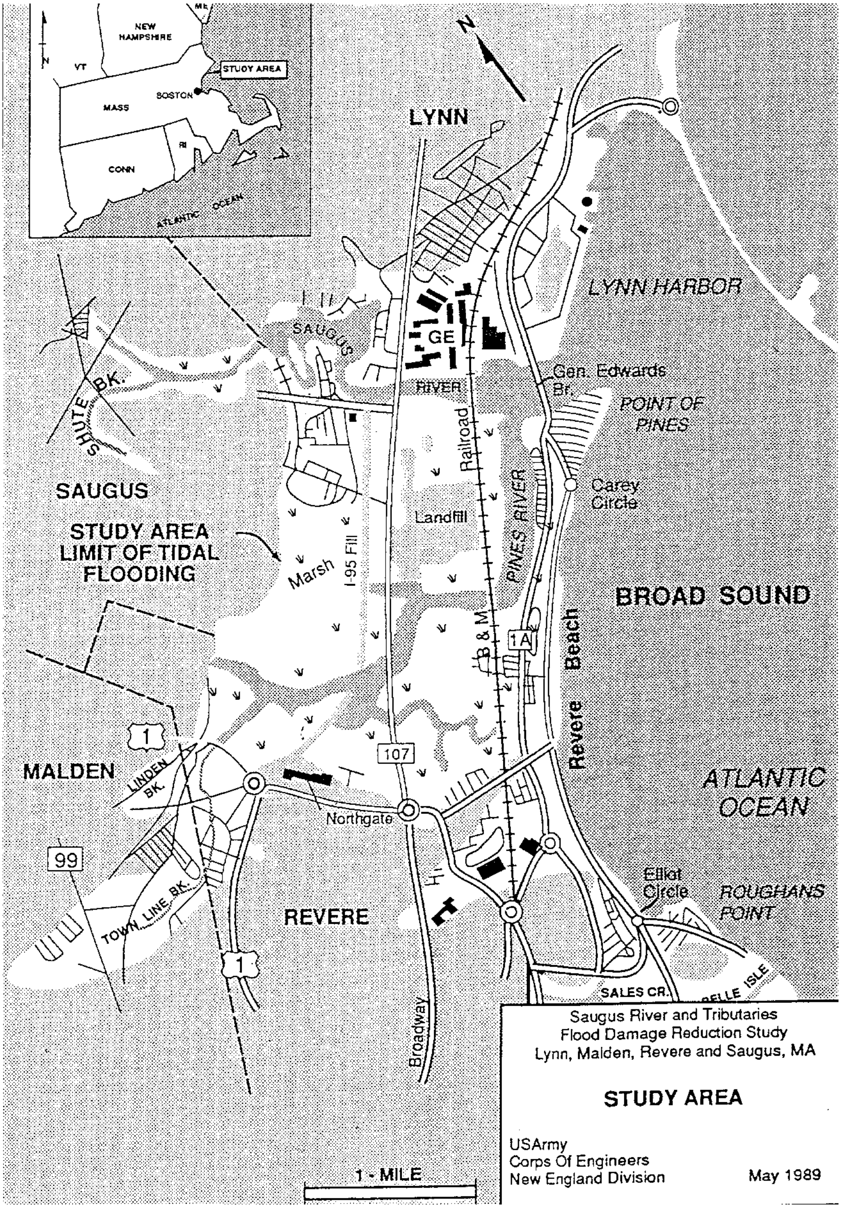


Fig. 1. Map of the Case Study Area. *Source:* U.S. Army Corps of Engineers (1990).

In this environment officials face the difficult task of developing policies to reduce the vulnerability of the area when uncertainty and surprise surround forecasts of future events, and uncertainty and ignorance cloud the understanding of the dynamics of vulnerability of the area to these storms. Further complicating the problem is that vulnerability is not spread evenly across the region. Each of the four communities has their own character, their own spatial configurations, and their own history of dealing with coastal storms (Fig. 1).

Revere is the home of the first public beach in the United States, Revere Beach, which still attracts many people from in and around the city of Boston on a hot summer day. However, during a major storm, approximately one-third of the city of Revere, including the Revere Beach Backshore behind Revere Beach and along the banks of the Pines River, is exposed to damages from wind, waves, and salt water flooding. One of the most vulnerable areas to coastal storms, Point of Pines, is a residential area located just north of Revere Beach along the Saugus River. This small peninsula is often affected by waves and wind, and flooding leaves many of its residents stranded (Manson et al., 2002). Largely residential, Revere also includes many high rise condominiums and retail office buildings along the beach, a light rail mass transit line, Boston & Maine Commuter Rail, Route 107, North Shore Road (1A), numerous marinas, and many small businesses (Clark et al., 1998)

Lynn, known as the City of Firsts, lies to the north of Revere and is more industrial and commercial than its neighbor. The Lynn Harbor shorefront includes the commercial and industrial district along Route 1-A, known as the Lynnway. This highway offers a direct access route for many of Lynn's businesses and industries and also serves a large number of North Shore commuters everyday. Major North Shore utilities including electric and gas distribution centers and a Regional Wastewater Treatment facility are located within the Lynn floodplain. A quick drive through the floodplain reveals many important facilities and roadways falling within close proximity to each other including the North Shore Community College, the General Electric River Works complex, MBTA facilities, West Lynn Creamery, Phillips Lighting, Norelco, numerous new and used car dealers, service stations, the Boston & Maine Commuter Rail, and the Salem Turnpike (Route 107).

Homes and businesses in the east portion of Saugus, located between the Saugus and Pines River marshes, suffer damages mostly due to frequent flooding. In the floodplains of the Upper Saugus and Shute Brook areas, town officials report that high tides cause drains to back up, flooding buildings in the center of town even though on higher ground. Homes often become flooded with a combination of high tides and runoff during major storms. The location of the first ironworks in the 1640s, Saugus has a long history of living and working with the river and the tides. During the operation of the ironworks, the Saugus River was used for transportation although the boats could only pass during high tide. In the 350 years since then, development has continued along the banks and marshes of the river despite the frequent flooding. In addition to many houses and businesses, the area now includes Route 107, the Boston and Maine Commuter Rail, most of the town's commercial navigation fleet and related facilities, elderly housing, a school, the Eastern Tool Company, RESCO Energy Systems, and several marinas.

The fourth community in the study area, Malden, lies further inland, southwest of Revere and Saugus. Only a small portion of Malden lies in the floodplain, the Town Line Brook and Linden Brook areas. This area is subject to flooding directly from the Pines River and from the backup drainage in the brooks during high tides. Most of Malden is on higher ground outside the floodplain leaving this area, immediately over the Malden–Revere boundary, a smaller problem for Malden than for the other towns. The Malden case makes clear that the spatial configurations of vulnerability do not follow municipal boundaries. Efforts in Malden to reduce vulnerability should be closely tied with efforts in the other towns, especially Revere, in a regional approach that crosses town lines and jurisdictions.

MEASURES OF VULNERABILITY

There are a number of policy options available to the public decision maker interested in reducing the vulnerability of an area to severe coastal storms and flooding. Floodplain protection measures can be either structural or non-structural in nature. Structural measures can include construction of a breakwater, a seawall, a dike, or a floodgate, beach restoration and nourishment, and sand dune development. Nonstructural measures can encompass a range of policy options aimed at preventing damages due to floods including floodproofing efforts, flood warning and evacuation, landuse management, flood insurance policies, and public acquisition of floodplain areas. It is the contention in this chapter that policy alternatives exist at different geographic scales. A decision to build a sea wall exists at a much different scale than a policy aimed at moving heating units out of basements that are susceptible to flooding. In addition, information exists at different scales and gives a different picture of reality at each scale. Changing the scale of the analysis can offer meaningful insights into the nature of the vulnerability. In older urban areas where much of the land supports mixed uses and high levels of diversity, a single public housing development or retirement home can be hidden among blocks of single family detached homes making the identification of the most effective policy alternatives for that area quite difficult. Should the city focus on education and outreach, evacuation assistance, or floodproofing and damage insurance?

The development of indices of vulnerability that we discuss in this chapter was part of a study conducted by the Marsh Institute at Clark University in the case study area. In Appendix A, we describe some of the sources of information used in that study; since we have relied on some information from the U.S. Army Corps of Engineers (1990) study, we have used the corresponding publically available data from the 1990 census. Our purpose in this chapter is to develop a method for constructing a relative index of vulnerability and to evaluate the properties and potential usefulness of that index. There has been a great deal of research devoted to the construction of indices to measure the vulnerability of populations to the hazards associated with global climate change (see, e.g., Clark et al., 1998; Cutter et al., 2000; Cutter, Boroff, & Shirley, 2003; Wu, Yarnel, & Fisher 2002; Rygel, O'Sullivan, & Yarnel, 2006; Eakin & Bojorquez-Tapia, 2008), all of which utilize some method of computing a weighted sum of vulnerability measures or attributes, for which the weights are subjectively obtained. To our knowledge, only the Clark et al. (1998) work employs DEA – and its method of objectively obtained weights – to construct their vulnerability indices. Another issue we investigate is how geographic scale affects the information provided by the vulnerability indices.

The scale of an index is tied directly to the scale of the information from which it is built. Therefore, we wanted to be able to work with data that is available at highly detailed levels of analysis, the block level, and uniformly measured throughout the case study area. This necessitated the use of Census data. For privacy reasons, only a subset of the full number of census variables is publically available at the block level. We chose to focus on six variables in this analysis. We are aware that they may not provide a fully comprehensive measure of vulnerability. The literature on vulnerability to climate change and the construction of vulnerability indices suggests numerous vulnerability dimensions or attributes for evaluation (see, e.g., Clark et al., 1998; Cutter et al., 2000, 2003; Wu et al., 2002; Rygel, O'Sullivan, & Yarnel, 2006). They do, however, represent a reasonable first measure of some of the underlying dimensions of vulnerability to this hazard and in this place. These six variables are number of persons of

Hispanic descent, number of persons of a minority race, number of persons over the age of 65, number of households with children under 18 and only one parent present, number of large (10 units or more) residential buildings, and mean rent level. The following discussion provides a brief justification for our choice of these variables.

Race and ethnicity are both identified in the literature as factors contributing to increased levels of vulnerability to natural hazards (Bolin & Bolton, 1986; Drabek & Key, 1984; Perry, Greene, & Mushkatel, 1983; Perry & Lindell, 1991).[1] Geographically isolated areas of Hispanics and recent Vietnamese immigrants in the case study area may require special outreach efforts including language translation and flood recovery assistance. The elderly is another subgroup within the population that may experience higher levels of vulnerability (Bolin, 1982; Bolin & Klenow, 1982; Drabek & Key, 1984; Quarentelli, 1991). Elderly persons can be less mobile and may require additional assistance with floodproofing efforts or during an evacuation. Similarly, income plays an important role in determining how quickly a person or a household will be able to recover from damage suffered during a storm (Clark et al, 1998; Bolin & Bolton, 1986; Bolin & Stanford, 1991; Drabek & Key, 1984; Perry & Lindell, 1991; Quarentelli, 1991). Although information on income levels cannot be obtained at the block level, we used mean rent level as a surrogate measure. The number of single-parent households was selected because parents dealing with the responsibilities of children on their own may have less flexibility, including child care options, to cope with the effects of a severe storm or flooding (Bolin & Bolton, 1986; Drabek & Key, 1984). The final factor measures the number of large (10 units or more) residential buildings. Large residences along the coastline require different policy actions pertaining to outreach and evacuation efforts.

The series of images (shown in Figs. 2–7) display the distribution of the six census variables at the census tract level in the four towns. Two tracts are identified in each figure, with A representing the tract with the lowest measure for that variable and B representing the highest.[2] In the next section, we describe and evaluate two methods for combining individual measures of vulnerability into composite indices.

CONSTRUCTING AN INDEX FOR COMPARISONS OF VULNERABILITY

Calculating and mapping each of the attributes that contribute to vulnerability can provide information useful in the formulation and implementation of

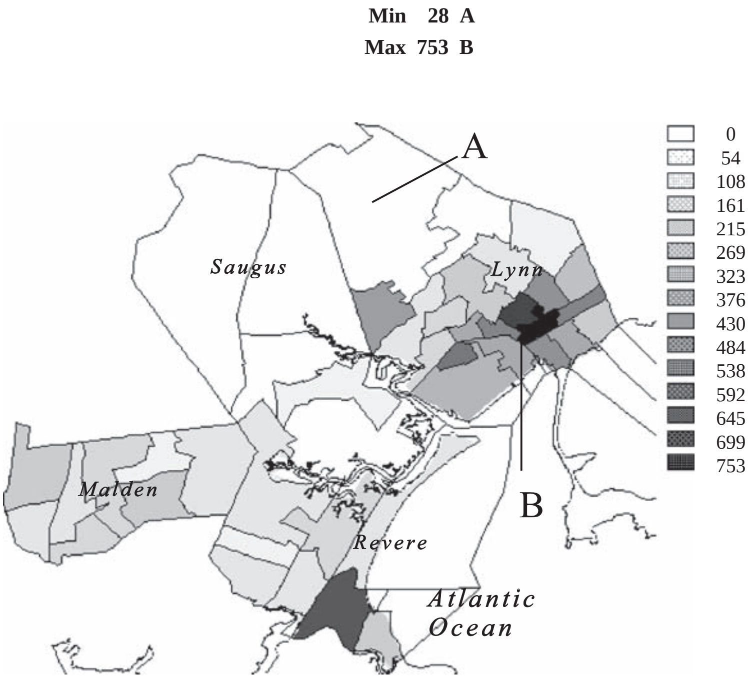


Fig. 2. Hispanic Population in Revere, Lynn, Malden, Saugus Census Tracts.

policies and programs designed to help avoid exposure or strengthen coping abilities. Evaluating these individual measures is important, but can become a complex task if there are a large number of attributes and spatial areas. In these cases, developing a measure, such as a composite index, that provides an overall assessment of the potential vulnerability of each area becomes important. There are a number of ways in which such indices can be developed (Ott, 1978). The most common approach is to create a weighted linear combination (or weighted product) of the attributes, the weighted average (or weighted geometric mean). In this section, we develop another approach for the construction of a relative composite vulnerability index by using DEA. We feel this approach complements the information provided by measures such as the weighted average and helps to overcome some of their shortcomings. This section describes how the DEA index is formulated and provides an illustrative example. In the following section, we discuss the results of using the two methods of index construction with the six variables measured in the case study area.

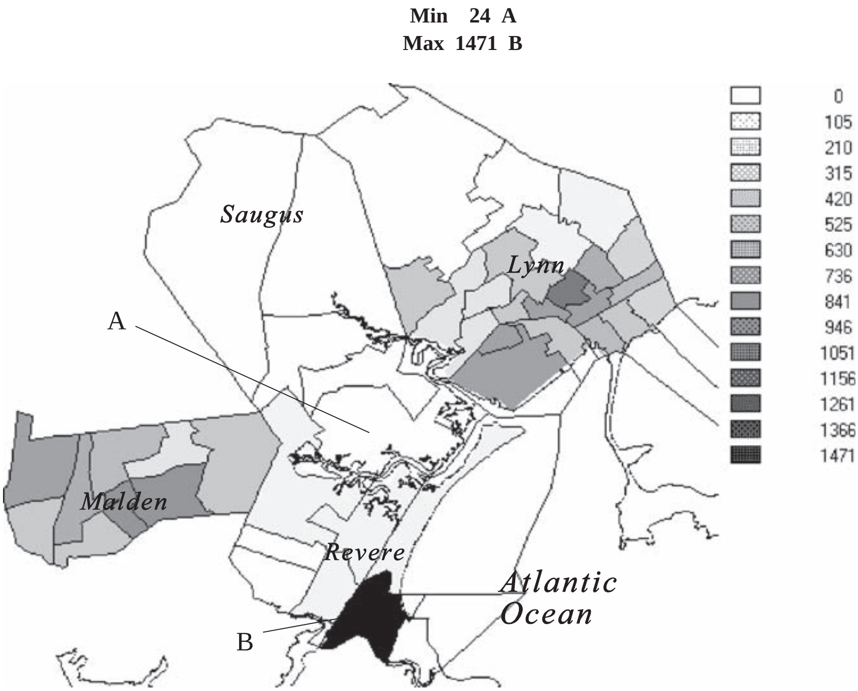


Fig. 3. Minority Populations in Revere, Lynn, Malden, Saugus Census Tracts.

An example of the generic form for the construction of the weighted average that is often used in the construction of vulnerability indices is provided in Eq. (1) below.

$$I_j = \sum_{i \in A} W_i M_{ij} \qquad \forall j \in J \tag{1}$$

where,

$I_j =$ The composite vulnerability index for geographic area j (census tracts, block, or block group) within the study area;
$W_i =$ The importance weight assigned to vulnerability attribute i;
$M_{ij} =$ The measure of vulnerability attribute i in geographic area j;
$A =$ The set of attributes that contribute to vulnerability; and
$J =$ The set of geographic areas that comprise the case study area (e.g., census tracts, block groups, or blocks).

Simple and direct, this type of combinatoric still has its problems. The importance weights W_i need to be obtained in some manner. One way to

Min 257 A
Max 1174 B

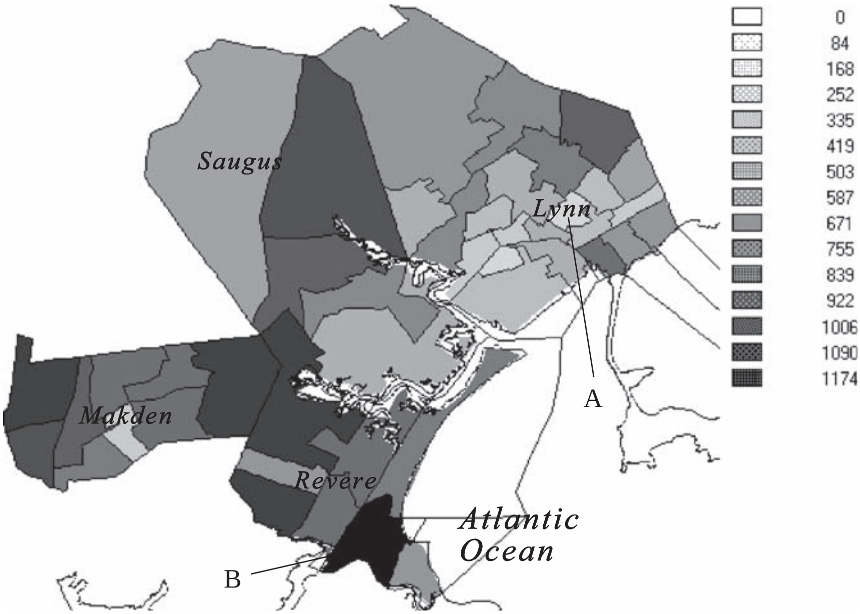


Fig. 4. Elderly Population in Revere, Lynn, Malden, Saugus Census Tracts.

obtain the weights would be to appeal to commonly accepted theory for assigning importance. For many types of decision problems, this option is not available. Another approach is to elicit preferences for the weights through some participatory mechanism involving either experts or relevant interest groups (Saaty, 1990; Korhonen, Moskowitz, & Wallenius, 1992; Yoon & Hwang, 1995; Eastman, Kyem, Toledano, & Jin, 1993; Eakin et al. 2008). This may lead to an unstable set of weights as participants change their minds or new interests are added to or dropped from the process. The weights also need to be adjusted to account for differences in the scales of measurement used to obtain M_{ij} for each attribute; changing a unit of measure will change the relative effect the importance weights have on the final index value. Another problem is that the weighted average can mask the effects of a single or an important subset of attributes.

We have used a DEA formulation, in addition to the weighted average, to construct composite vulnerability indices for the case study area. DEA has its theoretical and technical basis in operations research and the economic

Min 257 A
Max 405 B

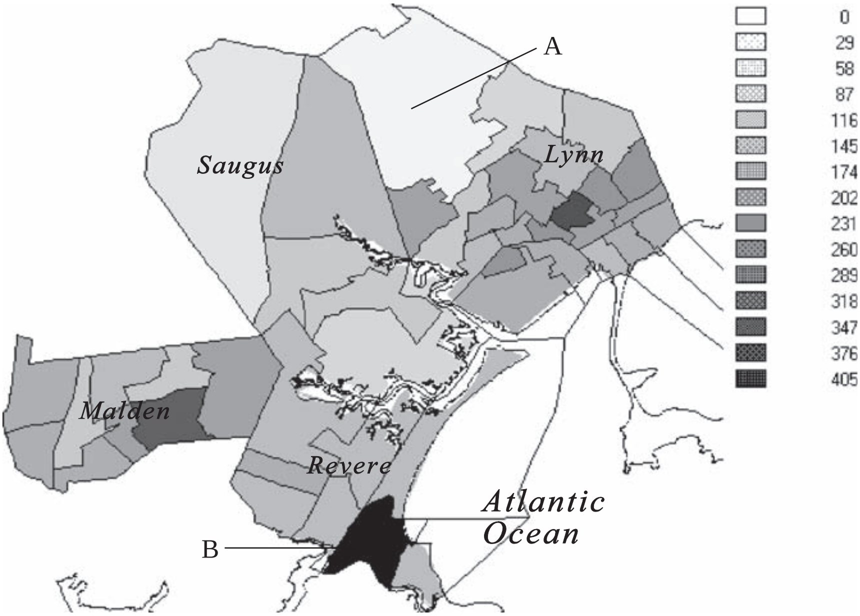


Fig. *5*. Single-Parent Households in Revere, Lynn, Malden, Saugus Census Tracts.

theory of production (Charnes, Cooper, & Rhodes, 1978; Farrell, 1957). DEA utilizes linear programming to produce measures of the relative efficiency of comparable decision-making units (DMUs) that employ multiple inputs and outputs. One dimension of the efficiency of an operation of any organization is the manner by which that organization selects and uses resources to produce its products; efficiency measures whether resources, equipment, and/or people are being put to good use. The more product produced for a given amount of resources the more efficient (i.e., less wasteful) is the operation. Mathematically, efficiency can be defined as the ratio of weighted outputs to weighted inputs:

$$E = \text{Efficiency} = \frac{\text{weighted sum of outputs}}{\text{weighted sum of inputs}}$$

The DEA approach objectively identifies the set of weights (all weights must be positive) that individually maximizes each DMU's efficiency while

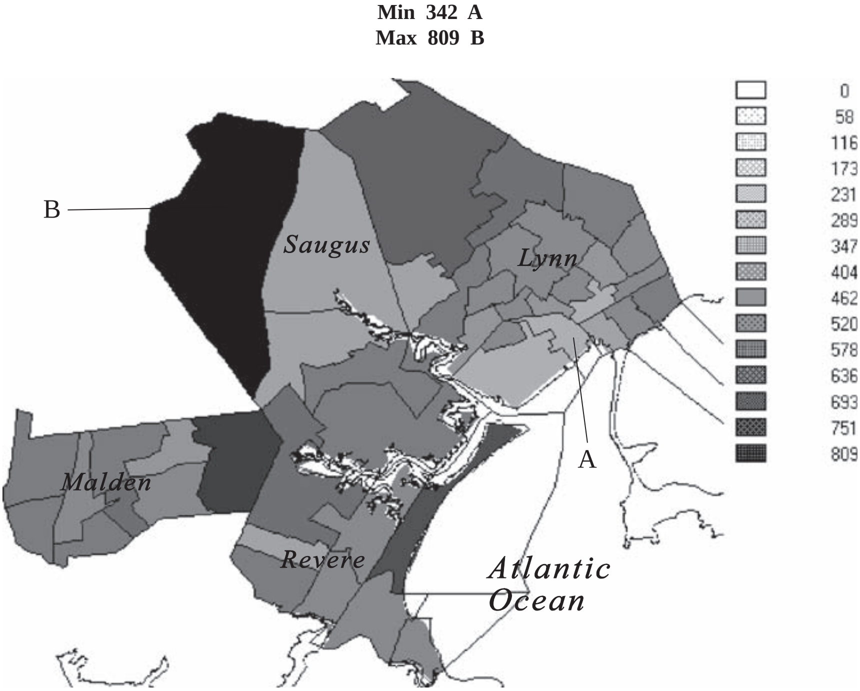


Fig. 6. Mean Rent Levels in Revere, Lynn, Malden, Saugus Census Tracts.

requiring the corresponding weighted ratios (i.e., using the same weights for all DMUs) of the other DMUs to be less than or equal to one. A DMU is considered relatively inefficient if its efficiency rating is less than one (i.e., $E<1$). The degree of inefficiency for a DMU is measured relative to a set of more efficient DMUs. However, a DMU identified as being efficient (i.e., $E = 1$) does not necessarily imply absolute efficiency. It is only relatively efficient as compared to the other DMUs that are being considered.

The DMU is the component of the organization that is being evaluated, (e.g., hospitals, departments, individuals, or water system size categories). DEA takes into account multiple inputs and outputs to produce a single aggregate measure of relative efficiency for each DMU. It requires only that the selected inputs and outputs be quantifiable. The technique can analyze these multiple inputs and outputs in their natural physical units without reducing or transforming them into some common measurement such as dollars.

The Charnes, Cooper, and Rhodes (CCR) DEA model (1978) is a linear fractional program that compares the ratio of weighted outputs to weighed

Min 0 A
Max 1784 B

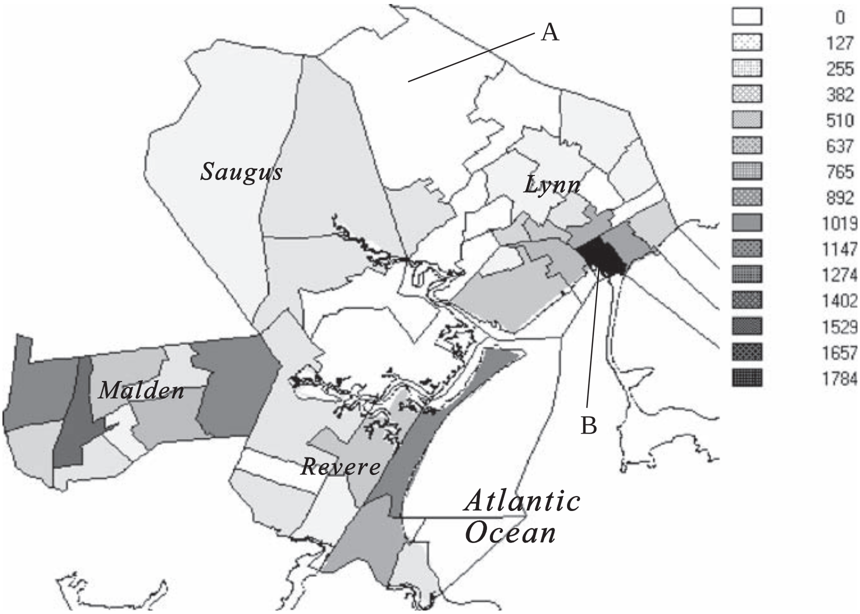


Fig. 7. Large Residential Buildings Revere, Lynn, Malden, Saugus Census Tracts.

inputs. The efficiency of the rth DMU, w_r, is obtained by solving the following linear fractional formulation:

$$\text{Maximize} \quad w_r = \frac{\sum_{j=1}^{j} u_j \mathrm{O}_{jr}}{\sum_{i=1}^{I} v_i I_{ir}}$$

s.t.

$$\frac{\sum_{j=1}^{j} u_j \mathrm{O}_{jk}}{\sum_{\mathrm{i}=1}^{\mathrm{I}} v_i I_{ik}} \quad \leq \forall\, k \tag{2}$$

$$u_{\mathrm{j}} \geq 0 \forall j,$$
$$v_{\mathrm{i}} \geq 0 \forall i$$

where,

$k = 1, \ldots, K$ DMU
$i = 1, \ldots, I$ Inputs
$j = 1, \ldots, J$ Outputs

Parameters

O_{jk} = amount of the *j*th output for the *k*th DMU
I_{ik} = amount of the *i*th input for the *k*th DMU

Decision variables

u_j = the weight assigned to the *j*th output and
v_i = the weight assigned to the *i*th input.

This initial CCR DEA model, (Eq. (2)), is a linear fractional program and is not truly a linear program and cannot be solved using the simplex method of linear programming. However, a few simple modifications will transform this initial formulation into a linear program. First, we assume that the denominator is equal to 1. We make this assumption into a constraint: $\sum_{i=1}^{I} v_i I_{ir} = 1$. The other modification is to multiply both sides of the equation for each constraint in Eq. (2) by the sum of the weighted inputs. This change linearizes each constraint in Eq. (2), to yield the following:

$$\sum_{j=1}^{J} u_j O_{jk} \leq \sum_{i=1}^{I} v_i I_{ik} \quad \forall k$$

Additionally, a special case called weakly efficient causes the DEA model to be modified in practice. A particular DMU is weakly efficient if it is actually inefficient, but when applying the DEA model, one or more of its weights, u_j or v_i, are equal to zero and it is incorrectly considered efficient. To avoid such situations, we add the restriction that the weights are required to be greater than some small number ε, where ε is a small infinitesimal value. The positivity of the weights, u_j and v_i, guarantees a weakly efficient DMU would not be found efficient. The modified linear CCR DEA formulation therefore becomes:

$$
\begin{aligned}
&\text{Maximize} \quad && w_r = \sum_{j=1}^{j} u_j \mathrm{O}_{jr} \\
&\text{s.t.} && \sum_{\mathrm{i}=1}^{\mathrm{I}} v_i I_{ir} = 1 \\
& && \sum_{j=1}^{j} u_j \mathrm{O}_{jk} - \sum_{\mathrm{i}=1}^{\mathrm{I}} v_i I_{ik} \leq 0 \quad \forall k = 1, \ldots, K \\
& && u_j, v_i \geq \varepsilon \quad \forall j, i
\end{aligned}
\tag{3}
$$

The CCR DEA formulation determines objectively the set of weights, u_r and v_i, that maximizes the efficiency of DMU$_k$, E_k. The constraints require the efficiency of each DMU, including the kth DMU, not exceed 1, and the weights, u_r and v_i, must be positive. A similar DEA formulation must be solved for each DMU. A DMU is considered relatively inefficient (i.e., $E_k < 1$) if it is possible to increase its outputs without increasing inputs or decrease its inputs without decreasing outputs. These efficiency ratings allow decision makers to identify which DMUs are in need of improvement and to what degree.

Since the Charnes, et al.'s 1978 publication, there have been thousands of theoretical contributions and practical applications in various fields using DEA (e.g., Desai, Ratick, & Shinnar, 2005; Emrouznedjad, Parker, & Tavares, 2008; Klimberg & Kern, 1992; Seiford, 1996). A unique variation of DEA used in this chapter is the concept of a unitary input (or no inputs) in the model. This unitary input DEA variation has been used to rank corporate performance based on financial ratios (Fernandez-Castro & Smith, 1994), evaluate public road services, (Odeck, 2005), and rank the performance of countries at the Olympic Games (De Mello, Angulo-Meza, & Da Silva, 2009). DEA has been applied to many diverse areas such as environment, health care, military operations, criminal courts, university departments, banks, electric utilities mining operations, manufacturing productivity, and railroad property evaluation (Emrouznedjad, 2009; Emrouznedjad et al., 2008; Gattoufi & Reisman 2004; Klimberg & Kern, 1992; Seiford, 1996; Tavares, 2002). More specifically, in the environmental area, DEA has been used to assess municipal solid waste collection (Courcelle, Kestemont, Tyteca, & Installé, 1998), manage protected forest regions (Kao, 2000), evaluate the pulp and paper industry (Hailu, 2003), measure eco-efficiency and cost-benefit analysis (Kuosmanen & Kortelainen, 2005), and evaluate environmental performance, (Ball, Färe, Grosskopf, & Nehring, 2001; Piot-Lepetit, Vermersch, & Weaver, 1997; Reinhard, Lovell, & Thijssen, 2000).

The mathematical formulation of DEA can be considered independent of the theory and can be used, as it is in this chapter, to construct scalar index values for the multidimensional concept of vulnerability (Clark et al., 1998; Cummings-Saxton, Ratick, & Desai, 1993; Cummings-Saxton, Ratick, Garriga, & Desai, 1994; Haynes, Ratick, & Cummings-Saxton, 1996). Adapting the general formulation for DEA (as shown in Eq. (3)) to create the vulnerability indices yields:

$$\text{Maximize } I_0 = \frac{\sum_{i \in A} W_{i0} M_{i0}}{\sum_{i \in AD} WD_{i0} MD_{i0}}$$

Subject to the constraints:

$$\frac{\sum_{i \in A} W_{i0} M_{ij}}{\sum_{i \in AD} WD_{i0} MD_{ij}} \leq 1 \quad \forall j \in J \tag{4}$$

$$W_{i0} \text{ and } WD_{i0} \geq \varepsilon \quad \forall i \in A \text{ or } AD$$

where:

I_0 = The composite vulnerability index for the geographic area (0) under consideration (census tracts, block, or block group) within the study region
W_{i0} = The weight assigned to vulnerability attribute i for geographic area 0
M_{i0} = The measure of vulnerability attribute i in geographic area 0
M_{ij} = The measure of vulnerability attribute i in geographic area j
WD_{i0} = The importance weight assigned to attribute i in area 0, where attribute i increases the coping ability of an area
MD_{ij} = The measure of attribute i in geographic area j, where attribute i increases the coping ability of geographic area j
AD = The set of attributes that increase coping ability in an area

Here, the DMUs are geographical areas. The DEA formulation shown above is used to find the weights W_{i0} and WD_{i0} for all attributes i, which will give the area under consideration (0) the highest possible composite vulnerability index score (constrained to be no larger than 1).

This score represents the ratio of the weighted linear combination of attributes that contribute to increased vulnerability, to the weighted linear combination of attributes that diminish vulnerability (or enhance coping ability). The constraints given in Eq. (4) also ensure that the composite vulnerability scores for all the other areas in the case study region, calculated with the weights W_{i0} and WD_{i0}, do not exceed 1. This optimization is repeated for each of the areas separately, each time finding weights to give that area its highest possible vulnerability score. The result provides a *relative* vulnerability ranking of the areas that comprise the case study region.

The following illustrative example demonstrates the construction and some of the properties of the DEA composite vulnerability index. For example, we consider 15 geographic areas in which two attributes that contribute to vulnerability have been measured (Table 1). Higher values of the attributes imply a more vulnerable area. Because we are only considering attributes that contribute to vulnerability (i.e., the attributes in set A), the denominators of the fractional programming problem are set to 1.

The DEA programming problem for finding the vulnerability index for area 11 can be written as:

$$\text{Maximize } I_{11} = W_{1,11} \times M_{1,11} + W_{2,11} \times M_{2,11}$$

Table 1. Example Problem Data and Results.

Geographic Area	Vulnerability Attribute 1	Vulnerability Attribute 2	Constraint Values using Weights for Area 11	DEA Vulnerability Index	Average
1	10	3	0.06	0.10	6.5
2	20	50	0.52	0.53	35.0
3	15	35	0.37	0.37	25.0
4	25	9	0.17	0.26	17.0
5	80	80	1.00	1.00	80.0
6	100	6	0.39	1.00	53.0
7	1	66	0.60	0.67	33.5
8	7	98	0.92	1.00	52.5
9	45	93	1.00	1.00	69.0
10	98	32	0.62	1.00	65.0
11	45	75	0.84	0.84	60.0
12	68	56	0.74	0.81	62.0
13	78	55	0.77	0.90	66.5
14	67	15	0.36	0.68	41.0
15	45	40	0.52	0.55	42.5

subject to the following constraints:

$$W_{1,11} \times M_{1,1} + W_{2,11} \times M_{2,1} \leq 1$$
$$W_{1,11} \times M_{1,2} + W_{2,11} \times M_{2,2} \leq 1$$
$$\vdots$$
$$W_{1,11} \times M_{1,15} + W_{2,11} \times M_{2,15} \leq 1$$

The solution gives weights $W_{1,11} = .0034$ and $W_{2,11} = .0091$. The values in the fourth column of Table 1 are the result of applying these weights to each of the areas in the constraints. The value of the weighted combination of attributes for area 11 is 0.84, which also represents the optimal value of the objective function. A higher vulnerability score for area 11 could not be obtained because increasing the weights would cause the constraint values for areas 5 and 9 to exceed 1. Likewise, if we are solving for the DEA vulnerability score for area 9, the weights associated with that solution are $W_{1,9} = .0013$ and $W_{2,9} = .0101$. When these weights are applied to each of the constraints, the value for the weighted combination of attributes for area 9 reaches the maximum of 1.00, and none of the constraints for the other areas exceed 1.

The DEA formulation was then solved for all 15 areas and the results presented in the fifth column of Table 1 and in Fig. 8. Positions of the areas in Fig. 8 are related to the amount of each attribute in each area (the DEA vulnerability score is given in parenthesis after the number for the region). Area 11, when compared to area 9, has the same value for attribute 1 but less of attribute 2, and area 5 (also with a composite index value of 1.00) has higher values of both attribute 1 and attribute 2 when compared to area 11. The DEA composite vulnerability index score of 0.84 reflects that comparison. The last column in Table 1 provides the average of the attributes for each area (each of the attributes is assigned an importance weight of 0.5). There seems to be general agreement between the average and the DEA scores for most areas. There is an exception for area 8, which has the highest value for attribute 2, but, one of lowest values for attribute 1. The average shows a mid-level value of 52.5, reflecting the masking effect of the low score. The DEA vulnerability index for the area, however, gives a relative score of 1.00 by virtue of the high value in attribute 2. (A similar situation occurs with area 6, which has a high value for attribute 1 but a low value for attribute 2.) This type of masking effect in the weighted average could be exacerbated as larger numbers of attributes or areas are considered.

The relative DEA composite vulnerability index scores would not be changed if we changed the "size" of the measurement (we use the word

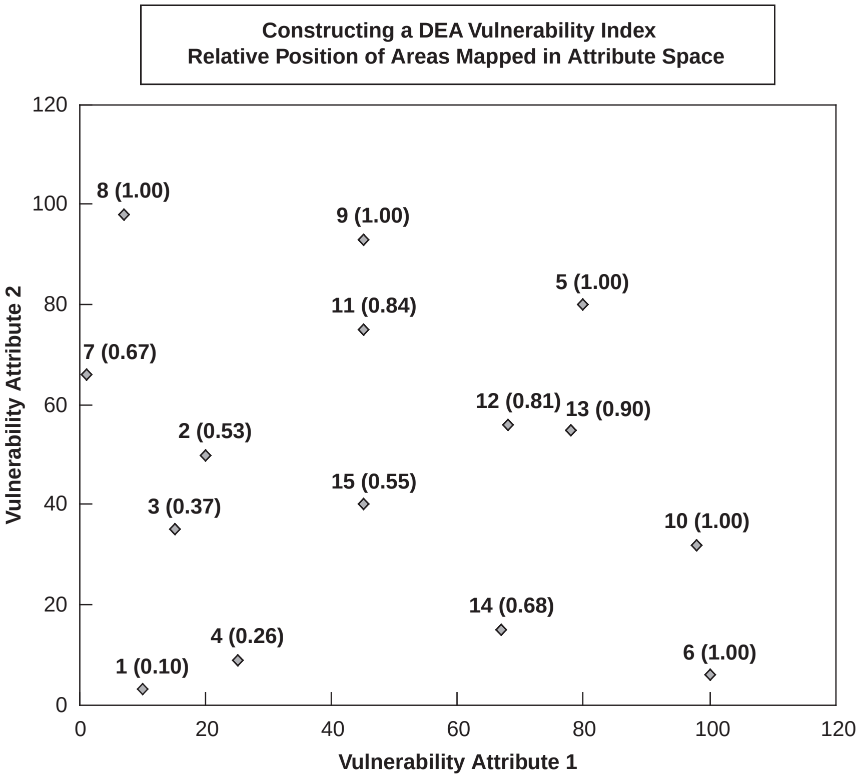


Fig. 8. Constructing a DEA Vulnerability Index.

"size" of a measure here instead of the more commonly used word "scale" to distinguish between the units in which an attribute is being measured and the geographic scale of measurement, which is discussed later in the chapter). This would negate any effect of pre-multiplying attribute values by importance or preference weights. The DEA index scores are a scalar representation of the relative position of the areas when mapped in attribute space, which would not be affected by re-sizing the axes.

The example helps to demonstrate how the DEA constructed index can complement the weighted average and help to address some of the problems in using a weighted average alone. The DEA index "weights" for an area are obtained objectively through the programming formulation and do not rely on any subjective assessment of importance. The DEA index values also can reflect the influence of one or a subset of attributes on the vulnerability score. As was demonstrated with area 8, one attribute having a relatively high value was enough to give that area a high vulnerability index score.

Changing units of measure, or applying importance weights, would not change the DEA index scores, which in some circumstances would be a strength of the procedure, but may be a drawback in others. There are a few approaches that have been proposed to address this issue. One is to create constraints on the weights called "assurance regions" (such as the weight for attribute 1 cannot be larger than that for attribute 2) (Thompson, Singleton, Thrall, & Smith, 1986). If the weights become too constrained, the advantages of the DEA index over the weighted average are diminished. Another approach is to develop separate DEA index weights for subsets of attributes that have similar importance values (called Multi-Objective DEA (MODEA), see Klimberg & Puddicombe, 1999). This reduces the number of items to be compared, but still leaves the issues of how these separate indices should be evaluated and perhaps combined. Another issue is that the DEA index values, as relative measures, rely on the specification of the region

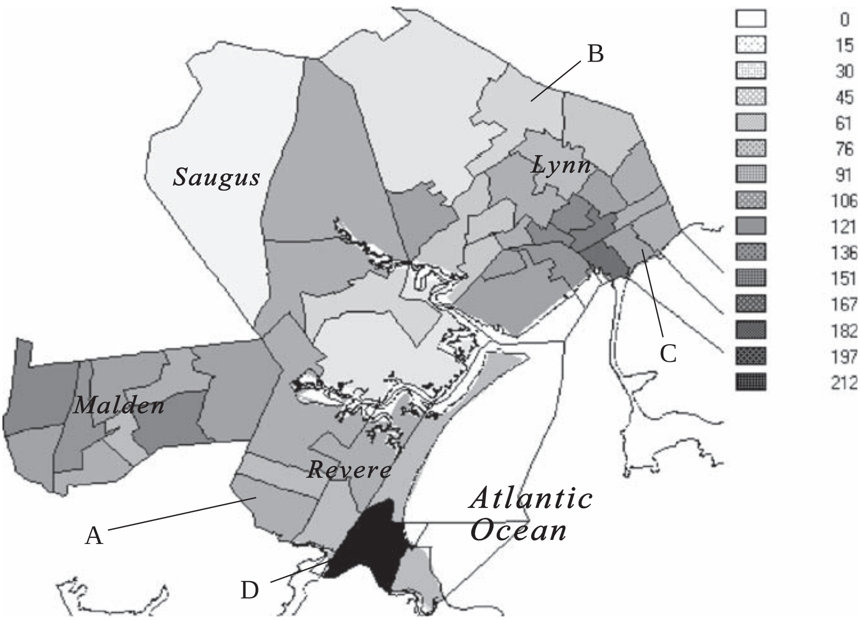
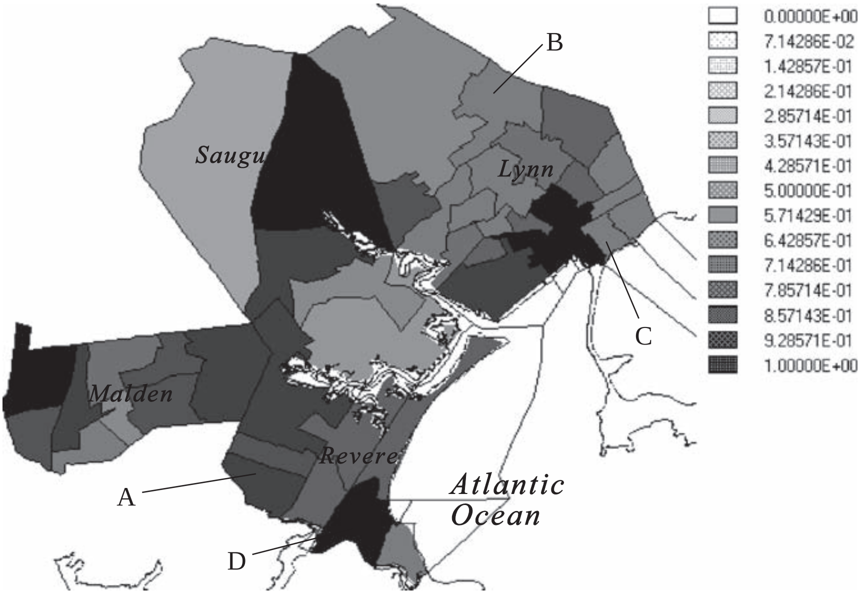


Fig. 9a. Weighted Average for Four Towns – Census Tracts.

under study. Adding areas to be evaluated or changing geographic scale may change the results. In the next section, we create, compare, and assess vulnerability indices for the case study area, which are developed using DEA and a weighted average.



Wavg	DEA		Hispanic	Minority	Over65	1Parent	10units	Rent
Mid	**High**	**A**	118	74	1105	159	302	565
Low	**Mid**	**B**	28	48	704	47	0	684
Mid	**Mid**	**C**	356	402	660	173	863	526
High	**High**	**D**	675	1471	1174	405	814	498
		Max	753	1471	1174	405	1784	342*
		Min	28	24	257	47	0	809*

* A low (high) mean rent indicates a high (low) level of vulnerability along this dimension.

Fig. 9b. DEA Scores for Four Towns – Census Tracts.

RESULTS OF THE ANALYSIS USING CENSUS DATA

Figs. 9(a) and (b) show the results of using the weighted average and the DEA formulation for creating indices of vulnerability for the four towns in the case study using the six variables measured at the census tract level. (Fig. 9(a) is the weighted average, Fig. 9(b) is the DEA index scores.) In the census tracts identified as C and D, the two techniques rate the locations similarly, a mid level of vulnerability and a high-level of vulnerability, respectively. However, in the tracts identified as A and B, the two techniques achieve different results. In both these cases, because DEA is maximizing, the DEA scores the tract as being more vulnerable than does the weighted average. In the case of B, the area rates low on all of the dimensions except for one. There is a relatively large elderly population in B that could be vulnerable to the impacts of a severe storm (see also Fig. 4, which shows the distribution of the elderly population). This information is obscured in the weighted average; the DEA technique highlights this one dimension of vulnerability and thus gives the tract a higher score (similarly with tract A). When analyzing policies aimed at mitigating vulnerability, it seems appropriate to consider the higher

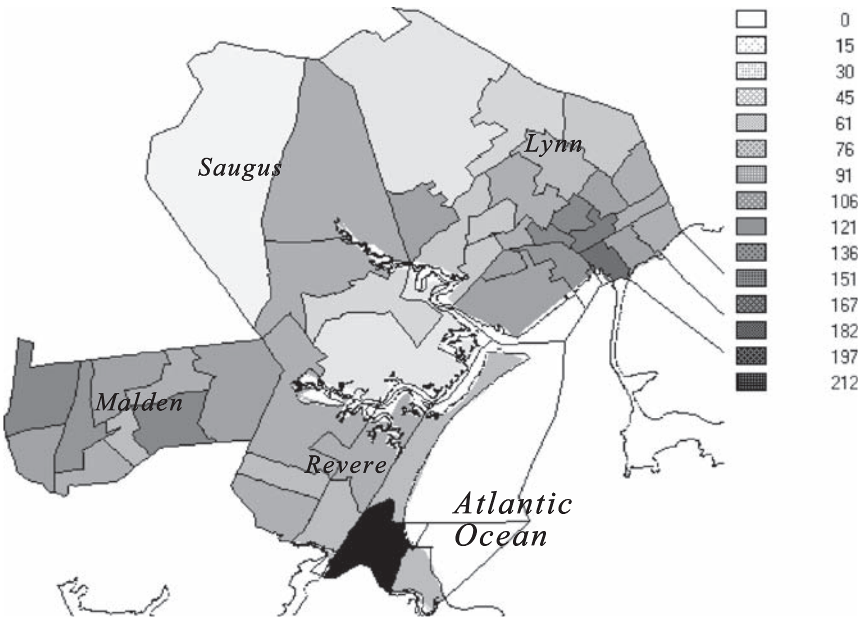


Fig. 10a. Weighted Average – Census Tracts.

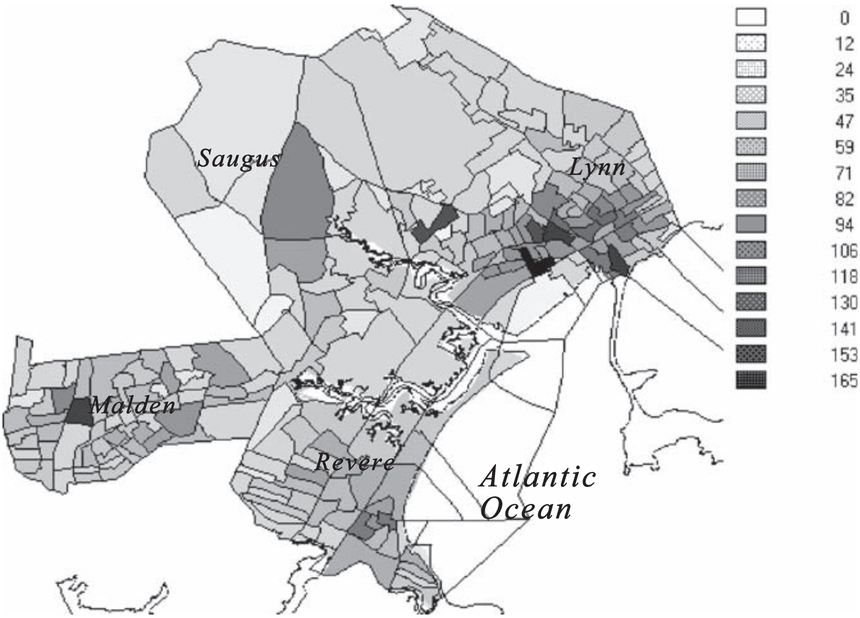


Fig. 10b. Weighted Average – Block Groups.

vulnerability values provided by the DEA index approach. The minimum and the maximum for each of the measures are included at the bottom of the table to facilitate such comparisons.

Figs. 10 and 11 illustrate the effect of changing the geographic scale of analysis on each of the techniques. (Figs. 10(a) and 10(b) show the results of the weighted average, and Figs. 11(a) and 11(b) show the results of the DEA.) The two images across the top of the figure are for the four towns at the level of the census tract. The bottom images show the four towns, but this time, broken down to the level of the block group. In each case, increasing the level of geographic detail changes the picture of vulnerability throughout the towns. There are fewer dark areas, representing high vulnerability, and they are now better specified. One can also clearly identify where highly vulnerability areas are adjacent to areas of low vulnerability – information that gets subsumed within the tract level. A situation where high vulnerability is more evenly spread throughout an area implies a different set of policy options than a situation where the decision maker needs to focus attention on an isolated pocket of high vulnerability within a surrounding area of relatively low vulnerability.

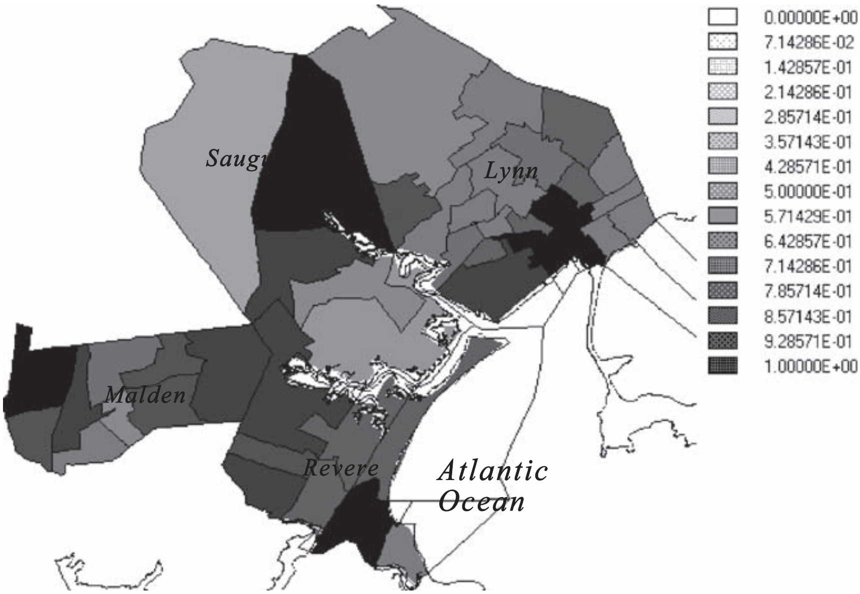


Fig. 11a. DEA Scores – Census Tracts.

The relationship between policy alternatives and geographic scale is further pronounced at the most refined level of detail – the census block level (Fig. 12). The patchwork pattern shown in Fig. 12 illustrates the high level of diversity among individual blocks in Revere. A single public housing unit or elderly residence, suggesting a higher level of vulnerability, may often be surrounded by blocks of single-family, middle-income homes. This diversity suggests the need to explore policies targeted at specific blocks or even specific buildings – not only policies addressing vulnerability across a wider area.

Fig. 12 illustrates how differential vulnerability can be hidden within the geographic scale of analysis. Consider Block A in Fig. 12, which falls within Block Group B in Figs. 13(a) (b) and Census Tract D in Figs. 9(a) (b). At each level, the area under question rates as highly vulnerable; however, as the scale is further refined, the area of high vulnerability becomes more targeted and smaller in spatial extent. The particular block shown in this example holds a large population of Vietnamese residents rating a high level of vulnerability because it is a minority population, living in multi-family homes, with a number of small children and elderly persons. This analysis reveals that special outreach efforts targeted at this particular population (such as language translation in public service announcements and

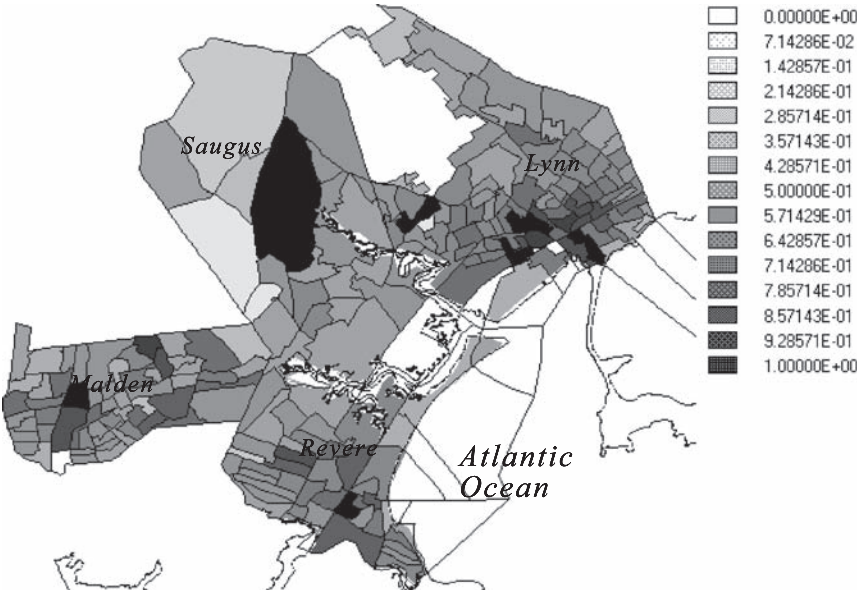


Fig. 11b. DEA Scores – Block Groups.

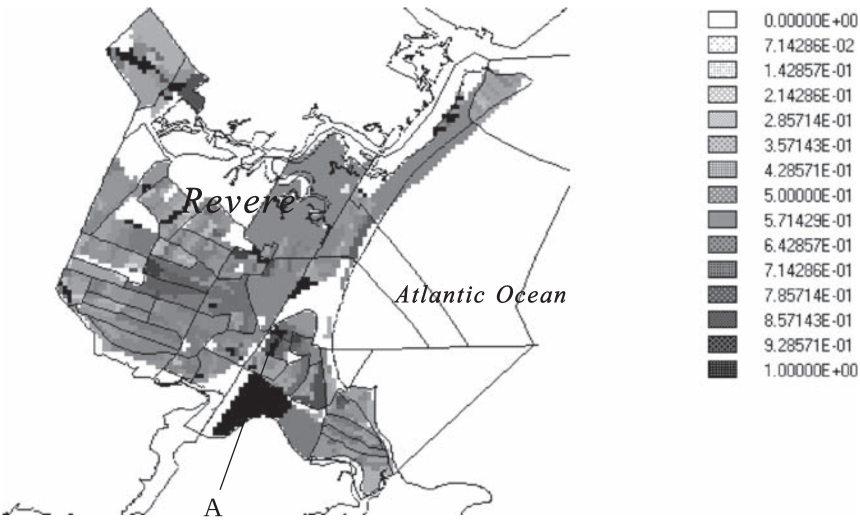


Fig. 12. DEA Scores – Census Block Level for Revere.

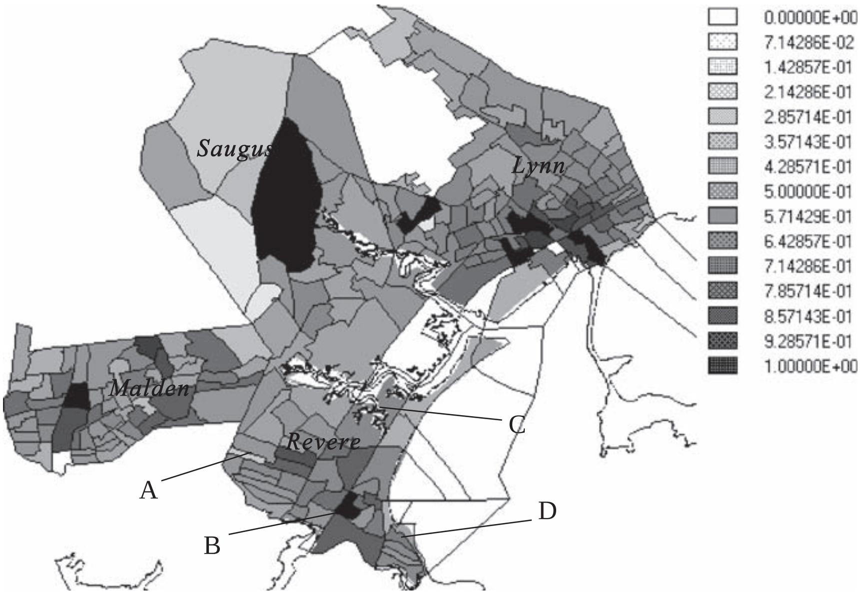
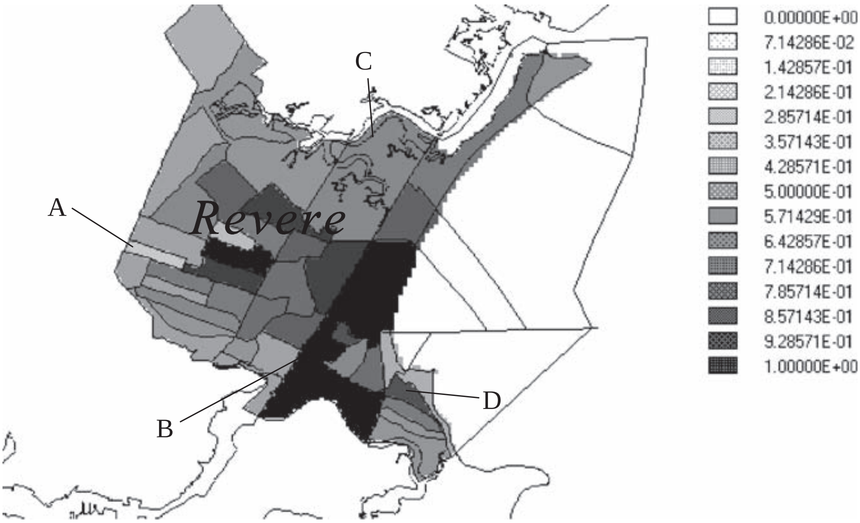


Fig. 13a. DEA Scores – Four Towns Block Group.

evacuation warnings, targeted flood mitigation assistance, or designated contact person(s) within city government) may provide the best alternatives for alleviating vulnerability.

Because DEA provides a relative index, its values are sensitive to the geographic scale of measurement – as is demonstrated above. The issue of geographic scale of measurement and the validity of the inferences that can be drawn from analyses done at different scales – related to the change of support problem (COSP) in geostatistics – has been well studied (e.g., Paez & Scott, 2004; Cressie, 1996); the caveats and suggested ways of overcoming these problems in this literature are pertinent in the development and use of a DEA vulnerability index.

Drawing out one final point about the use of DEA in index construction, Figs. 13(a) and (b) illustrate the sensitivity of relative measures, such as DEA scores, to the set of locations included in the analysis. The image in Fig. 13(a) shows the DEA scores when all four towns are included in the analysis; the image in Fig. 13(b) shows the DEA scores for Revere, when just the block groups in Revere are evaluated. Block Group D is seen as having only a mid-level vulnerability score when compared with block



Wavg	DEA		Hispanic	Minority	Over65	1Parent	10units	Rent
Low	**Low**	**A**	5	0	95	16	1	751
High	**High**	**B**	106	570	96	101	51	533
Mid	**Mid**	**C**	15	11	205	28	0	532
Mid	**High**	**D**	116	27	190	46	175	475
		4 Town Max	333	570	756	138	1207	208*
		Revere Max	166	570	395	101	611	308*

* A low (high) mean rent indicates a high (low) level of vulnerability along this dimension.

Fig. 13b. DEA Scores – Revere Block Groups.

groups in all four towns, but for Revere, it is one of the most vulnerable areas within the city. This comparison distinguishes what is *relatively* vulnerable in Revere, and how that same area rates when compared to the surrounding towns. Which outlook is most appropriate would depend on the decision maker's jurisdiction and on the available policy options.

CONCLUSIONS

In this chapter, we have discussed the development of composite indices for measuring and mapping the vulnerability of populations to coastal hazards associated with severe storms. The relative composite index, developed using DEA, was shown to complement the information provided by other measures, specifically the weighted average. The results of the analysis demonstrate how DEA can help to overcome some of the problems of using a weighted average type measure, including (1) the need to assign importance weights that fix the "substitutability" of the attributes comprising vulnerability and, (2) the masking of the influence of one (or a subset) of the attributes contributing to a measure of vulnerability.

Through the application of these indices to four costal towns along Boston's North Shore, we have attempted to demonstrate their potential usefulness to policy formulation and implementation. The towns in the case study area have a long history of severe coastal storms and many years experience in determining policies aimed at preventing and mitigating wind, wave, and flood damages. A number of policy alternatives exist to reduce the vulnerability of an area to coastal hazards, generally falling into two categories: structural or non-structural measures. Structural measures usually try to reduce the potential exposure of populations to flood effects through the construction of breakwaters, seawalls, dikes, or floodgates; beach restoration and nourishment; and sand dune development. Construction of floodgates was one of the major options considered by the Corps of Engineers for the Saugus-Pines watershed (U.S. Army Corps of Engineers, 1990). Non-structural measures, which generally attempt to address the resistance and resilience of populations to the consequences of flood hazards, include floodproofing efforts, flood warning and evacuation, land-use management, flood insurance policies, and public acquisition of floodplain areas. The scale of information needed to assess the necessity of building floodgates is different from that needed to help design evacuation programs. Because policy alternatives are designed for and implemented at different geographic scales, we have explored the types of information provided by the vulnerability indices at various scales of analysis.

We realize that an array of definitions and measures of vulnerability exist. For the reasons explained above, we have chosen a selected subset of population attributes to develop the composite index and to assess its properties. In limiting the scope of this chapter, we have not included an evaluation or mapping of the differential exposure of the population in the case study area to storm hazards of varying intensities and duration.

Including the differential impacts of exposure across the case study area would be necessary for a more complete assessment of vulnerability. Given these limitations, we have attempted to establish the usefulness of the DEA relative index measure and of spatial representations of composite indices to the design of policies aimed at reducing vulnerability.

NOTES

1. It should be noted that some research shows that strong family and support networks operating within ethnic or minority enclaves may not increase the vulnerability of those areas but may actually operate to strengthen people's coping abilities (Perry, Greene, & Mushkatel, 1983; Perry & Lindell, 1991) again demonstrating the differing and often conflicting views about what makes a population vulnerable.
2. We chose to use raw numbers in these displays and in the index construction exercise shown in this article. In some cases, it may be preferable to normalize the data by using, for example, percentages. However, we felt that from a policy point of view being able to distinguish between a tract that has 400 elderly people out of a total of 800 and a tract that has 4 elderly people out of 8 would be more informative than knowing that in both locations 50% of the population is 65 years old or over.

ACKNOWLEDGMENTS

We are greatly indebted to many individuals for their input and contributions over the course of several years. We thank our colleagues in the Graduate School of Geography, the George Perkins Marsh Institute, and Clark Labs whose contributions have been invaluable: George Clark, Kirsten Dow, J. Ronald Eastman, Sri Emani, Dominic Golding, Hong Jiang, Weigen Jin, Jeanne Kasperson, Roger Kasperson, Susi Moser, William B. Myer, and Harry Schwarz. We also like to acknowledge the National Oceanic and Atmospheric Administration (NOAA) for funding parts of this research.

REFERENCES

Ball, V. E., Färe, R., Grosskopf, S., & Nehring, R. (2001). Productivity of the U.S. agricultural sector: The case of undesirable outputs. Chap. 13. In: C. R. Hulten, E. R. Dean & M. J. Harper (Eds), *New developments in productivity analysis*. Chicago: University of Chicago Press for the National Bureau of Economic Research.

Barth, M. C., & Titus, J. G. (Eds). (1984). *Greenhouse effect and sea level rise*. New York: Van Nostrand Reinhold.

Bolin, R. (1982). *Long-term family recovery from disaster* (University of Colorado. Chapters 1, 2, and 8). Monograph #36. Boulder, CO: Institute of Behavioral Science.

Bolin, R., & Bolton, P. (1986). *Race, religion, and ethnicity in disaster recovery*. Program on Environment and Behavior Monograph #42. Boulder, CO: Institute of Behavioral Science, University of Colorado.

Bolin, R., & Klenow, D. J. (1982). Response of the elderly to disaster: An age-stratified analysis. *International Journal of Aging and Human Development*, *16*, 283–295.

Bolin, R., & Stanford, L. (1991). Shelter, housing, and recovery: A comparison of U.S. disasters. *in Disasters*, *15*(1), 24–34.

Brooks Nick, W., Adger, N., & Kelly, P. M. (2005). The determinants of vulnerability and adaptive capacity at the national level and the implications for adaptation. *Global Environmental Change*, *15*, 151–163.

Cannon, T. (1994). Vulnerability analysis and the explanation of 'natural' disasters. In: A. Varley (Ed.), *Disasters, Development, and Environment* (pp. 13–30). New York: Wiley.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, *2*, 429–444.

Clark, G. E., Moser, S. C., Ratick, S. J., Dow, K., Meyer, W. B., Emani, S., Jin, W., Kasperson, J. X., Kasperson, R. E., & Schwarz, H. E. (1998). Assessing the vulnerability of coastal communities to extreme storms: The case of Revere, MA, USA. *Mitigation and Adaption Strategies for Global Change*, *3*, 59–82.

Courcell, C., Kestemont, M. P., Tyteca, D., & Installé, M. (1998). Assessing the economic and environmental performance of municipal solid waste collection and sorting programmes. *Waste Management and Research*, *16*, 253–263.

Cressie, N. (1996). Change of support and the modifiable areal unit problem. *Geographical Systems*, *3*, 159–180.

Cummings-Saxton, J., Ratick, S. J., & Desai, A. (1993). *Pollution prevention frontiers (PPF): An approach to measuring pollution prevention progress*. Cambridge, MA: Industrial Economics, Inc.

Cummings-Saxton, J., Ratick, S. J., Garriga, H. M., & Desai, A. (1994). *Pollution prevention frontiers (PPF) and other approaches to pollution prevention assessment: Comparisons based on New Jersey materials accounting data*. Cambridge, MA: Industrial Economics, Inc.

Cutter, S. L. (1996). *Vulnerability to environmental hazards*. Hazard's Research Lab Discussion Paper Series.Department of Geography, University of South Carolina.

Cutter, S. L., Boroff, B. J., & Shirley, W. L. (2003). Social vulnerability to environmental hazards. *Social Science Quarterly*, *84*(2), 242–260.

Cutter, S. L., Mitchell, J. T., & Scott, M. C. (2000). Revealing the vulnerability of people and places: A case study of Georgetown county, South Carolina. *Annals of the Association of American Geographers*, *90*(4), 713–737.

De Mello, J. C., Angulo-Meza, L., & Da Silva, B. P. B. (2009). A ranking for the Olympic Games with unitary input DEA models. *IMA Journal of Management Mathematics*, *20*, 201–211.

Desai, A., Ratick, S. J., & Shinnar, A. (2005). Data envelopment analysis with stochastic variations in data. *Socio-Economic Planning Sciences*, *39*, 147–164.

Dow, K. (1992). Exploring differences in our common future(s): The meaning of vulnerability to global environmental change. *Geoforum*, *23*(3), 417–436.

Drabek, T. E. (1986). *Human systems response to disasters: An inventory of sociological findings*. New York: Springer-Verlag.

Drabek, T. E., & Key, W. H. (1984). *Conquering disaster: Family recovery and long-term consequences*. New York: Irvington Publishers.

Eakin, H., & Bojorquez-Tapia, L. A. (2008). Insights into the composition of household vulnerability from multicriteria decision analysis. *Global Environmental Change*, *18*, 112–127.

Eastman, J. R., Kyem, R. A. K., Toledano, J., & Jin, w. (1993). *GIS and decision making*. Geneva: UNITAR.

Emrouznedjad, A. (2009). Data Envelopment Analysis home page. Available at http://www.DEAzone.com

Emrouznedjad, A., Parker, B. R., & Tavares, G. (2008). Evaluation of research in efficiency and productivity: A survey and analysis of the first 30 years of scholarly literature in DEA. *Socio-Economic Planning Sciences*, *44*, 151–157.

Farrell, M. J. (1957). The measure of productive efficiency. *Journal of Regional Statistical Society*, Series A, *120*(Pt. 3), 253–281.

Fernandez-Castro, A., & Smith, P. (1994). Towards a general non-parametric model of corporate performance. *Omega*, *22*(30), 237–249.

Gattoufi, S. O. M., & Reisman, A. (2004). A taxonomy for data envelopment analysis. *Socio-Economic Planning Sciences*, *38*(2–3), 141–158.

Hailu, A. (2003). Pollution abatement and productivity performance of regional Canadian pulp and paper industries. *Journal of Forest Economics*, *9*, 5–25.

Haynes, K. E., Ratick, S. J., & Cummings-Saxton, J. (1996). Toward a pollution prevention abatement monitoring policy: Measurement, model mechanics, and data requirements. *The Environmental Professional*.

Kao, C. (2000). Data envelopment analysis in resource allocation: An application to forest management. *International Journal of Systems Science*, *31*(9), 1059–1066.

Klimberg, R. K., & Kern, D. (1992). *Understanding data envelopment analysis (DEA)*. Boston University School of Management Working Paper, pp. 92–44.

Klimberg, R. K., & Puddicombe, M. (1999). A multiple objective approach to data envelopment analysis. *Advances in Mathematical Programming and Financial Planning*, Greenwich: JAI Press, 5(1999), 201–232.

Korhonen, P., Moskowitz, H., & Wallenius, J. (1992). Multiple criteria decision support: A review. *European Journal of Operational Research*, *62*, 361–375.

Kuosmanen, T., & Kortelainen, M. (2005). Measuring eco-efficiency of production with data envelopment analysis. *Journal of Industrial Ecology*, *9*(4), 59–72.

Manson, S. M., Ratick, S. J., & Solow, A. R. (2002). Decision making and uncertainty: Bayesian analysis of potential flood heights. *Geographical Analysis*, *34*(2), 112–129.

Mitchell, J. K. (1989). *Risk assessment of environmental change*. Working Paper no. 13. Environment and Policy Institute, East West Center, Honolulu.

Odeck, J. (2005). Evaluating target achievement in the public sector: An application of a rare non-parametric DEA Malmquist indices. *Journal of Applied Economics*, *8*(1), 171–190.

Ott, W. R. (1978). *Environmental indices – theory and practice*. Ann Arbor, MI: Ann Arbor Science.

Paez, A., & Scott, D. M. (2004). Spatial statistics for urban analysis: A review of techniques with examples. *GeoJournal*, *61*, 53–67.

Perry, R. W., Greene, M., & Mushkatel, A. (1983). *American minority citizens in disaster*. Seattle, Washington: Battelle.

Perry, R. W., & Lindell, M. K. (1991). The effects of ethnicity on evacuation decision making. *International Journal of Mass Emergencies and Disasters*, *9*(1), 47–68.

Piot-Lepetit, I., Vermersch, D., & Weaver, R. D. (1997). Agriculture's environmental externalities: DEA evidence for French agriculture. *Applied Economics*, *29*, 331–338.

Quarentelli, E. L. (1991). *Disaster assistance and socioeconomic recovery at the individual and household level: Some observations*. Preliminary Paper No. 171. Newark, Delaware: University of Delaware, Disaster Research Center.

Ratick, S. J., Baughman, M., Dow, K., Du, W., Moran, S., Solow, A., & Broadus, J. (1992). *Coastal hazards and global environmental change*. Report to the Institute for Water Resources, Water Resources Support Center, U.S. Army Corps of Engineers.

Reinhard, S., Lovell, C. A. K., & Thijssen, G. J. (2000). Environmental efficiency with multiple environmentally detrimental variables; estimated with SFA and DEA. *European Journal of Operational Research*, *121*, 287–303.

Rygel, L., O'Sullivan, D., & Yarnel, B. (2006). A method for constructing a social vulnerability index: An application to hurricane storm surges in a developed country. *Mitigation and Adaptation Strategies for Global Change*, *11*, 741–764.

Saaty, T. L. (1990). How to make a decision: The analytic hierarchy process. *European Journal of Operational Research*, *48*, 9–26.

Schneider, S. H., & Chen, R. S. (1980). Carbon dioxide warming and coastline flooding: Physical frameworks and climatic impacts. *Annual Review of Energy*, *5*, 107–140.

Seiford, L. M. (1996). Data envelopment analysis: The evaluation of the state of the art (1978–1995). *The Journal of Productivity Analysis*, *9*, 99–137.

Solow, A. R., & Ratick, S. J. (1994). Conditional simulation and the value of information. In: R. Simitrakopoulos (Ed.), *Geostatistics for the next Century* (pp. 209–217). Norwell, MA: Kluwer Academic Publishers.

Susman, T., O'Keefe, T., & Wisner, T. (1983). Global disasters: A radical interpretation. In: K. Hewitt (Ed.), *Interpretations of Calamity* (pp. 263–283). Boston: Allen and Unwin.

Tavares, G. (2002). A bibliography of data envelopment analysis (1978–2001). RUTCOR, Rutgers University. Available at http://rutcor.rutgers.edu/pub/rrr/reports2002.1_2002.pdf

Thompson, R. G., Singleton, F. D., Thrall, R. M., & Smith, B. A. (1986). Comparative site evaluations for locating a high-energy physics lab in Texas. *Interfaces*, *16*, 35–49.

Timmerman, P. (1981). *Vulnerability, resilience, and the collapse of society*. Toronto: Institute for Environmental Studies, University of Toronto.

U.S. Army Corps of Engineers. (1990). *Flood damage reduction: Main report – section I.* Feasibility Report and Final Environmental Impact Statement/Report. Water Resources Investigation. New England Division, U.S. Army Corps of Engineers.

Wu, S. Y., Yarnel, B., & Fisher, A. (2002). Vulnerability of coastal communities to sea level rise, a case study of Cape May County, New Jersey, USA. *Climate Research*, *22*, 255–270.

Yoon, K. P., & Hwang, C. (1995). *Multiple attribute decision making: An introduction*. London: Sage.

APPENDIX A. SOURCES OF INFORMATION

Table A1 lists federal, state/regional, and local sources of publicly available data relating to vulnerability to severe coastal storms on the North Shore. The U.S. Army Corps of Engineers (Corps) is the principal source of information on exposure including estimates of damages for different flood levels, hydrology studies, and the Flood Damage Reduction Report (U.S. Army Corps of Engineers, 1990). The Federal and Massachusetts Emergency Management Administrations (FEMA and MEMA) also provide records of reported damage in the four towns due to storms, and local and area newspapers offer descriptive reports on the extent and magnitude of events. In addition, the National Flood Insurance Program and the Small Business Administration keep records of claims that were made following the storm, and public officials and departments (police, fire, civil defense, MBTA, public utilities, etc.) relate stories of emergency response efforts and storm impacts.

Information on the population's coping ability, resilience, and resistance, has been collected from a variety of sources. The 1990 U.S. Census is a principal source of data on both the housing and the built environment in the towns and on the socioeconomic characteristics of the population. Variables pulled out of the census identify age and type of housing, owner occupied housing, race and ethnicity, family structure, income, age, education, disabilities, transients, and immigrants, among other factors. Key individuals in the communities are also good sources of information on the character, people, history, and social norms or way of life in each of the towns. Stories from residents themselves provide an understanding of the type and range of experiences faced by people living and working in the area.

The available sources of information identified in Table A1 include data that take a number of different formats. Some of the data, such as the census data or the Corps hydrology reports, is quantitative and standardized. This is the type of data that the project relies on most heavily in the construction of the vulnerability indices. The information found in the more qualitative formats, newspaper articles, interviews, second-hand stories, or accounts transform the impersonal numbers and percentages into the real experiences of real people. Other data sources fall between these two extremes. For instance, insurance claims, damage reports, and records of assistance are more formal in their reporting methods but lack the standardized spatial configurations offered by census units – census tracts, block groups, blocks levels. The study uses these other sources of information to inform the overall understanding of vulnerability; however, the vulnerability indices are constructed solely from census data and damage estimates. Further research will investigate ways in which the more qualitative types of information, as available at various geographic scales, can be combined with the available quantitative data in the actual construction of the measures of vulnerability.

Table A1. Sources of Information.

Federal	State/Regional	Local
1990 U.S. Census	Massachusetts Emergency Management Administration (MEMA)	Local newspapers
Toxic Releases Inventory (TRI) Database		
Federal Emergency Management Administration (FEMA)	Metropolitan Boston Transportation Authority (MBTA)	Town and Housing Authority
Platt Maps	Red Cross (regional)	Town maps
National Flood Insurance Program	Massachusetts Department of Mental Health	Fire and Police Departments
U.S. Army Corps of Engineers (Corps):	Massachusetts DUA	City planners
Flood Damage Reduction Report	Public Utilities	Civil Defense Director
	Massachusetts Department of Labor	Area churches and Salvation Army
Stage Damage Curves		
Hydrology Studies	Boston Globe (newspaper)	Public Works Departments
Small Business Administration	Area Companies and Large Employers	Area Residents and Business Owners

SECTION C
DATA ENVELOPMENT ANALYSIS (DEA) APPLICATIONS II

PERFORMANCE EVALUATION OF UNIVERSITIES FROM THE STUDENTS' PERSPECTIVE

Andreas Kleine and Regina Schlindwein

ABSTRACT

DEA is a favored method to investigate the efficiency of institutions that provide educational services. We measure the efficiency of German universities especially from the students' perspective. Since 1998, the Centrum für Hochschulentwicklung *(CHE) evaluates German universities annually. The CHE ranking consists of three ranking groups for different indicators, but they do not create a hierarchy of the universities. Thus, a differentiation of the universities ranked in the same group is not possible. Based on the CHE data set, especially the surveys among students, we evaluate teaching performance from the students' point of view using data envelopment analysis (DEA). DEA enables us to identify departments that – in the students' perspective – are efficient in the sense that they provide high quality of education. As a method for performance evaluation, we apply a DEA bootstrap approach. By the use of this approach, we incorporate stochastic influences in the data and derive confidence intervals for the efficiency. Based on data generated by the bootstrap procedure, we are able to identify stochastic efficient*

Financial Modeling Applications and Data Envelopment Applications
Applications of Management Science, Volume 13, 181–198

ISSN: 0276-8976/doi:10.1108/S0276-8976(2009)0000013012

departments. These universities serve as a benchmark to improve teaching performance.

1. INTRODUCTION

Universities intend to offer an excellent education and high quality research but measuring provided performance is a difficult task. This article deals with the performance evaluation of German universities focussing on the education of students and their qualification for the job market. Sarrico, Hogan, Dyson, and Athanassopoulos (1997) identify three classes of stakeholders interested in the universities' teaching performance.

(i) The government that represents the society wants universities to improve their teaching efficiency as they provide financial resources from tax money. It is necessary to increase the number of students as there is an increasing demand for qualified university graduates from the job market. In Germany, this is an important aspect as the proportion of the working population with university degrees is below OECD average (Al-Fahham, 2008).
(ii) The universities themselves belong to the second group. They are interested to offer high quality of education to their students. A good reputation in teaching will attract more and especially high-potential students. Since the introduction of tuition fees in some German Federal States, this aspect is of increasing importance. Competition amongst universities for the best students is higher than ever before.
(iii) The third group of stakeholders is the potential students, which need information for their choice of an appropriate university. Besides tuition fees, cost of living, and other factors, they are interested in high-quality education. In the following, we will focus on the perspective of these potential students.

Since 1998, the *Centrum für Hochschulentwicklung* (CHE) annually publishes a ranking of German universities. The primary goal of the ranking is to provide information on subjects and universities to prospective students. Unlike other rankings, the CHE ranking does not aggregate all criteria to achieve an overall hierarchy. It places the universities for different criteria into one of three groups: Top Group, Middle Group, and Bottom Group. This approach does not allow for any differentiation of universities or university departments in the same group.

In this study, we evaluate the performance of German university departments with a data envelopment analysis (DEA) approach. With the DEA model, we are able to identify the teaching efficiency of university departments and to set apart inefficient departments. For the university's perspective, the DEA model also provides reference universities that serve as a benchmark with respect to an improvement of the universities' teaching performance. However, a disadvantage of DEA is that it is an estimation technique relying on extreme points. Consequently, the estimates are subject to uncertainty due to sampling variation. To achieve more reliable results of the performance evaluation, we use the smoothed homogenous bootstrap procedure proposed by Simar and Wilson (1998) and subsequent publications. The DEA bootstrap approach considers stochastic influences in the data set and offers the possibility for statistical inference. Applying the bootstrap procedure enables estimating the empirical distribution functions of efficiency scores. We use these distribution functions to determine the first-order stochastic efficient universities. This allows for finding efficient references for inefficient universities.

DEA is an appropriate method for the evaluation of performance in education as it becomes apparent in numerous applications in this field. Already in its first application, DEA was used for the measurement of performance in education when Rhodes (1978) evaluated the program follow-through in public schools in the United States. It was the preparatory work for the development of the DEA method formalized by Charnes, Cooper, and Rhodes (1978). Since then, there have been many DEA applications in the field of education ranging from performance measurement in elementary schools over high schools to MBA programs. Most applications were made in Great Britain, the United States, and Australia. With regard to the performance measurement in universities, there are two main fields of investigation: education and research. An overview of DEA applications to educational institutions may be found in Gilles (2005) and Worthington (2001). There are also some DEA applications to German universities (Al-Fahham, 2008; Dyckhoff, Rassenhövel, Gilles, & Schmitz, 2005; Gutierrez, 2005; Warning, 2004).

This chapter is structured as follows. In Section 2, we give a brief description of the DEA method and its statistical properties. Furthermore, we introduce the DEA bootstrap procedure and a model for the first-order stochastic dominance (FSD). In Section 3, we apply the described methods for the evaluation of teaching performance of business administration departments in German universities. We describe the data used, the model specifications, and discuss the empirical results. Finally, conclusions are summarized and areas worth further investigation are identified.

2. METHODS

2.1. Data Envelopment Analysis and Its Statistical Properties

DEA is a method for performance measurement of decision-making units (DMUs) based on historical data. It implements a concept of relative efficiency, that is, a DMU is compared to best practices observed (Färe, Grosskopf, & Lovell, 1985; Cooper, Seiford, & Tone, 2007). The underlying production is characterized by a transformation of inputs into outputs. The set of production possibilities can be described as follows:

$$\Psi = \{(\mathbf{x}, \mathbf{y}) \in \mathbb{R}_+^{p+q} | \mathbf{x} \text{ can produce } \mathbf{y}\} \tag{1}$$

with $\mathbf{x} \in \mathbb{R}_+^p$ as a vector of inputs and $\mathbf{y} \in \mathbb{R}_+^q$ as a vector of outputs. In the following, we present an output-oriented view, as we will use an output-oriented model in the application to university departments.

Mathematically production possibilities are described by the production correspondence. In the output space, this is the set of outputs $\mathbf{y}$ that can be produced by inputs $\mathbf{x}$:

$$\mathbf{Y}(\mathbf{x}) = \{\mathbf{y} \in \mathbb{R}_+^q | (\mathbf{x}, \mathbf{y}) \in \Psi\} \tag{2}$$

The DMUs are characterized by their observed input output data $(\mathbf{x}_i, \mathbf{y}_i), i = 1, \ldots, n$. For determining the efficiency of DMU_0 $(0 \in \{1, \ldots, n\})$, we compare its production $(\mathbf{x}_0, \mathbf{y}_0)$ with a best practice production, that is, a production on the efficient frontier of the production possibilities. The Farrell efficiency boundary for the output space is

$$\partial\mathbf{Y}(\mathbf{x}) = \{\mathbf{y} | \mathbf{y} \in \mathbf{Y}(\mathbf{x}), \theta\mathbf{y} \notin \mathbf{Y}(\mathbf{x}), \forall\theta > 1\} \tag{3}$$

All productions on this boundary are efficient as it is not possible to produce a higher quantity of outputs with the given input. The Farrell output measure of efficiency of a DMU_0's production $(\mathbf{x}_0, \mathbf{y}_0)$ is defined as

$$\theta_0 = \theta(\mathbf{x}_0, \mathbf{y}_0) = \sup\{\theta | \theta\mathbf{y}_0 \in \mathbf{Y}(\mathbf{x})\} \tag{4}$$

where θ_0 measures the relative distance of DMU_0's production from the efficient frontier $\partial\mathbf{Y}(\mathbf{x})$. It is the largest possible radial augmentation of all outputs of DMU_0 within the set of production possibilities Ψ at a given level of inputs $\mathbf{x}_0$. That means, to become output efficient, DMU_0 should augment its outputs proportionally by θ_0. In the case of the output-oriented efficiency measure, a university department is efficient if its optimal value of θ_0 equals 1, and it is not efficient if the optimal value of θ_0 is greater than 1.

In general, the set of production possibilities Ψ is unknown and so are $\mathbf{Y}(\mathbf{x})$ and $\partial\mathbf{Y}(\mathbf{x})$. Only the observed input output data of the DMUs $\mathcal{X}_n = \{(\mathbf{x}_i, \mathbf{y}_i) : i = 1, \ldots, n\}$ is available. These observations are used to estimate the unknown quantities, and so, we have the estimators $\widehat{\Psi}, \hat{\mathbf{Y}}(\mathbf{x})$ and $\widehat{\partial\mathbf{Y}}(\mathbf{x})$. Thus, it is possible to estimate the efficiency of a DMU_0 by $\widehat{\theta_0}$. DEA is a popular nonparametric estimator for the set of production possibilities. It is based on assumptions regarding the process of how the observed data $\mathcal{X}_n$ is generated, the so-called *Data Generating Process* (DGP). Simar and Wilson (2000a) make the following assumptions regarding DGP $\mathcal{P}$: Ψ contains all observed productions, meaning the probability that a realization belongs to Ψ is 1, $\mathrm{Prob}\{(\mathbf{x}_i, \mathbf{y}_i) \in \Psi\} = 1$, and the set of production possibilities is convex. Observations in $\mathcal{X}_n$ are realizations of iid random variables on Ψ with the probability density function $f(\mathbf{x}, \mathbf{y})$. The density $f(\mathbf{x}, \mathbf{y})$ is strictly positive along the efficient frontier of Ψ as with an increasing sample size the probability that Ψ contains efficient productions is approaching unity. There is smoothness along the true frontier, that is, for all $(\mathbf{x}, \mathbf{y})$ in the interior of Ψ, $\theta(\mathbf{x}, \mathbf{y})$ is differentiable in both its arguments.

Under these assumptions, the DEA estimator for the set of production possibilities is

$$\widehat{\Psi} = \left\{ (\mathbf{x}, \mathbf{y}) \in \mathbb{R}_+^{p+q} \middle| \mathbf{y} \leq \sum_{i=1}^{n} \lambda_i \mathbf{y}_i, \mathbf{x} \geq \sum_{i=1}^{n} \lambda_i \mathbf{x}_i, \sum_{i=1}^{n} \lambda_i = 1, \lambda_i \geq 0, \forall i = 1, ..., n \right\} \tag{5}$$

for a production technology with variable returns of scale (VRS). Hence, the DEA estimator for technical output efficiency of DMU_0 is

$$\widehat{\theta_0} = \max \left\{ \theta \middle| \theta \mathbf{y} \leq \sum_{i=1}^{n} \lambda_i \mathbf{y}_i, \mathbf{x} \geq \sum_{i=1}^{n} \lambda_i \mathbf{x}_i, \sum_{i=1}^{n} \lambda_i = 1, \lambda_i \geq 0, \forall i = 1, ..., n \right\} \tag{6}$$

The DEA efficiency estimator $\widehat{\theta}$ measures the relative distance from the efficient frontier of $\widehat{\Psi}$. We know that according to its construction, $\widehat{\Psi} \subseteq \Psi$; consequently, the estimated efficiency is biased. That means that $\widehat{\theta} \leq \theta$ as the radial distance of a DMU from the efficient frontier of the estimated set of production possibilities $\widehat{\Psi}$ is smaller than its distance from the efficient frontier of the true production possibilities Ψ. Nevertheless, the biased estimation of efficiency is consistent, that is, $\widehat{\theta}$ converges toward the true efficiency θ as the sample size increases. For discussion of consistency and sampling distribution, see Simar and Wilson (2000a). Gijbels, Mammen, Park, and Simar (1999) show that it is possible to determine the bias ($bias = \theta - \widehat{\theta}$) and to correct the DEA efficiency estimator. Therefore, Simar and Wilson (1998) proposed a general methodology for bootstrapping in nonparametric frontier models, which is described in the following section.

2.2. Bootstrapping in the DEA Model

The bootstrapping method offers the possibility to make inferences without concrete knowledge of the underlying distribution. The idea is to approximate the unknown distribution of the estimators by repeatedly sampling with replacement from the original data set. In case of the DEA model, the distribution of the DEA efficiency estimator is approximated by a simulation of the DGP. The basis is the observed input output data. By repeatedly sampling and replacement of data, we create a pseudo sample and determine the DEA estimators for this simulated sample. We can repeat this process B times to have B pseudo samples and thus compute B bootstrap estimates $\widehat{\theta}_0^*$ for DMU_0. The empirical distribution of these bootstrap estimates is an approximation of the sampling distribution of DMU_0's efficiency estimate.

Suppose that $\widehat{\mathcal{P}}(\mathcal{X}_n)$ is a consistent estimator of the DGP $\mathcal{P}$. Then, we have the following relations for the efficiency scores

$$(\widehat{\theta}_0 - \widehat{\theta}_0^*)|\widehat{\mathcal{P}} \sim (\theta_0 - \widehat{\theta}_0)|\mathcal{P} \tag{7}$$

The DEA bootstrap procedure simulates the DGP with a Monte Carlo approximation. The error of the approximation depends partly on the value of B, that is, for $B \to \infty$, the error decreases with the number of replications B. Furthermore, the quality of the approximation increases with the sample size, $n \to \infty$.

As this so-called *naive bootstrap* yields inconsistent results, Simar and Wilson (1998) present a smoothed bootstrap procedure. In this procedure, the pseudo samples are not generated directly from the original input output data. The pseudo samples are rather generated by resampling the DEA efficiency estimators $\widehat{\theta}_i$. The algorithm for the *homogeneous smooth bootstrap procedure* that consistently generates the bootstrap efficiencies $\widehat{\theta}_i^*$ from a kernel density estimate as presented by Daraio and Simar (2007) can be summarized as follows:

1. For each DMU$_i$ with input output data ($\mathbf{x}_i$, $\mathbf{y}_i$), $i = 1, \ldots, n$, compute the DEA estimator $\widehat{\theta}_i$.
2. Reflect the n DEA estimators about unity and add the reflected points $(2 - \widehat{\theta}_i, \mathbf{x}_i, \mathbf{y}_i)$ to the original data $(\widehat{\theta}_i, \mathbf{x}_i, \mathbf{y}_i)$. This is necessary to achieve consistency. Determine the bandwidth parameter h for the multivariate setting as in Simar and Wilson (2000b).
3. Draw uniformly with replacement n bootstrap values β_i^* from the set of $2n$ original and reflected DEA scores.
4. Use the following random generator to smooth and perturb the bootstrap sample $\beta_i^*, i = 1, ..., n$ from the kernel density:

$$\tilde{\theta}_i = \beta_i^* + h\in_i, \quad \text{for} \quad i = 1, ..., n \tag{8}$$

   where h is the bandwidth parameter and $\in_i$ is a random draw from the standard normal distribution.
5. Correct the bootstrap values for the mean and the variance of the smoothed values by

$$\theta_i^{**} = \bar{\beta}^* + \frac{1}{\sqrt{1 + h^2/\widehat{\sigma}^2}}(\tilde{\theta}_i - \bar{\beta}^*) \quad \text{for} \quad i = 1, ..., n \tag{9}$$

   with $\bar{\beta}^* = \frac{1}{n}\sum_{i=1}^{n} \beta_i^*$ and $\widehat{\sigma}^2$ is the sample variance of the bootstrap sample $\tilde{\theta}_i, i = 1, ..., n$.
6. Reflect the values smaller than 1 to come back to measures greater than one by

$$\theta_i^* = \begin{cases} 2 - \theta_i^{**} & \text{for } \theta_i^{**} < 1 \\ \theta_i^{**} & \text{otherwise} \end{cases}$$

7. Use the smoothed bootstrap sample $\theta_{ib}^*, i = 1, ..., n$ of bootstrap iteration b to generate the new data set $\mathcal{X}_b^* = \{(\mathbf{x}_i^*, \mathbf{y}_{ib}^*) : i = 1, ..., n\}$ with $\mathbf{x}_i^* = \mathbf{x}_i$ and $\mathbf{y}_{ib}^* = \left(\widehat{\theta}_i/\theta_{ib}^*\right)\mathbf{y}_i$.
8. Compute the bootstrap efficiency estimator $\widehat{\theta_{ib}^*}$ by solving the DEA model as described above but using the new data set $\mathcal{X}_b^*$.
9. Repeat steps 3–8 B times for each DMU$_i$, $i = 1, \ldots, n$ to provide a set of bootstrap estimators $\widehat{\theta_{ib}^*}, b = 1, ..., B$.

The bootstrap procedure generates an empirical distribution of the bootstrap efficiency scores $\widehat{\theta_i^*}$ for each university department (DMU_i). The empirical bootstrap distribution allows us to estimate the bias of the DEA efficiency estimator, which is used to correct the efficiency estimator, the empirical distribution function, and confidence intervals for the true efficiency.

2.3. Statistical Inference

Just as the DEA estimator, the bootstrap estimator $\widehat{\theta^*}$ is biased by construction. By definition, we have *bias* $(\widehat{\theta_0}) = \theta_0 - E(\widehat{\theta_0})$. The bootstrap *bias* estimate of the DEA efficiency estimator $\widehat{\theta_0}$ is analogously defined as $\widehat{bias}(\widehat{\theta_0}) = \widehat{\theta_0} - \frac{1}{B}\sum_{b=1}^{B}\widehat{\theta_{0b}^*}$. A bias-corrected estimator of θ_0 is obtained by computing

$$\widehat{\widehat{\theta_0}} = \widehat{\theta_0} + \widehat{bias}(\widehat{\theta_0}) = 2\widehat{\theta_0} - \frac{1}{B}\sum_{b=1}^{B}\widehat{\theta_{0b}^*} \tag{10}$$

As the bias correction introduces additional noise and could have a higher mean square error than the DEA efficiency estimator, the bias correction should not be used unless (Efron & Tibshirani, 1993)

$$\widehat{\sigma}^2 < \frac{1}{3}\left[\widehat{bias}(\widehat{\theta_0})\right]^2 \tag{11}$$

with the estimate for the variance of $\widehat{\theta_0}$

$$\widehat{\sigma}^2 = \frac{1}{B}\sum_{b=1}^{B}\left[\widehat{\theta_{0b}^*} - \frac{1}{B}\sum_{b=1}^{B}\widehat{\theta_{0b}^*}\right]^2 \tag{12}$$

We use the bootstrap estimators $\widehat{\theta_{0b}^*}, b = 1, \ldots, B$ to determine a corrected empirical distribution function of the true efficiency. In accordance to bias correction (Eq. (10)), we calculate a bias-corrected efficiency estimator of DMU_0 for each bootstrap iteration b as follows

$$\widehat{\widehat{\theta}}_{0b} = \widehat{\theta_0} + \widehat{bias}_{0b} = \widehat{\theta_0} + (\widehat{\theta_0} - \widehat{\theta_{0b}^*}) = 2\widehat{\theta_0} - \widehat{\theta_{0b}^*} \tag{13}$$

For the estimation of the confidence intervals of the true efficiency, it is necessary to sort the values $\widehat{\widehat{\theta}}_{0b}, b = 1, \ldots, B$ for DMU_0 in an increasing order and then delete $(\alpha/2 \times 100)$–percent of the elements of either end of the list.

2.4. Stochastic Dominance

The estimation of the empirical distribution function of DEA efficiencies allows us to test the DMUs for stochastic dominance. We use the concept of FSD introduced by Hanoch and Levy (1969) to compare the DMUs. In case of an output-oriented DEA approach, a high-quality university department is characterized by small estimators, that is, the smaller a bootstrap estimator the better a department. A department is referred to as FSD-efficient, if we do not find another reference that has a higher probability for any arbitrary bootstrap level.

Let $F_{\theta_0}(s) := \text{Prob}\{\theta_0 \leq s\}$ the empirical distribution function of DMU_0's bootstrap efficiencies, then DMU_0 is FSD-efficient, if there does not exist a reference with estimator θ':

$$F_{\theta'}(s) \geq F_{\theta_0}(s),\ \forall s \in \mathbb{R} \text{ and } F_{\theta'}(s) > F_{\theta_0}(s),\ \exists s \in \mathbb{R} \tag{14}$$

meaning that the empirical distribution of a dominating reference is never below the empirical distribution of a dominated DMU. The efficiency probability of a dominating DMU is in all cases higher or at least equal.

Using discrete probability functions, Marx (2003) proofed that a comparison of each jump is a necessary and sufficient condition for FSD efficiency. Thus in case of bootstrapping, we have to compare bootstrap estimators $\widehat{\widehat{\theta}}_{ib}$ for each b. Kuosmanen (2004) introduced a mixed integer linear program to verify FSD efficiency. In contrast to Kuosmanen and to reduce complexity, we initially size all efficiency estimators for each DMU_i (w.l.g):

$$\widehat{\widehat{\theta}}_{i\langle 1\rangle} \leq \widehat{\widehat{\theta}}_{i\langle 2\rangle} \leq \ldots \leq \widehat{\widehat{\theta}}_{i\langle B\rangle} \quad \text{for } i = 1, ..., n \tag{15}$$

It is easier this way to check FSD efficiency by use of a linear program. Following a well-known idea of Wendell and Lee (1977), we introduce a model that is quite similar to DEA programs. Variables ξ_b measure possible reductions for the bootstrap estimator of each scenario $b = 1, \ldots, B$ and linear program (Eq. (16)) determines a reference with a maximal expected reduction:

$$z_0 = \max\left\{ \frac{1}{B}\sum_{b=1}^{B} \xi_b \;\middle|\; \begin{array}{l} \sum_{i=1}^{n} \widehat{\widehat{\theta}}_{i\langle b\rangle}\lambda_i \leq (1-\xi_b)\widehat{\widehat{\theta}}_{0\langle b\rangle}, \xi_b \geq 0, \text{for } b = 1, ..., B \\ \sum_{i=1}^{n} \lambda_i = 1, \lambda_i \geq 0, \text{for } i = 1, ..., n \end{array} \right\} \tag{16}$$

The optimal solution of Eq. (16) indicates whether DMU_0 is FSD-efficient or not. FSD efficiency is measured by: $\widehat{z_0} = 1/(1 - z_0)$. In accordance to estimator $\widehat{\theta_0}$, a DMU_0 is FSD-efficient if $\widehat{z_0} = 1$ and DMU_0 is dominated if $\widehat{z_0} > 1$. If we are interested in a comparison with only observed DMUs, we have to add $\lambda_i \in \{0,1\}$ for each DMU_i applying a so-called free disposal hull technology (fdh).

If a DMU is first-order dominated, we find a reference that is better with respect to the empirical distribution. Classical bootstrap methods do not provide any information of reference DMUs. Using linear program (Eq. (16)) and fdh technology however, each first-order dominated DMU can be compared directly with a corresponding efficient one. Moreover, a calculation of FSD efficiency for each DMU reveals the significance of a FSD-efficient DMU. In addition, the more frequently a university department serves as a reference the better is its performance.

3. APPLICATION AND RESULTS

3.1. Data and Model Specification

The present study uses data from the CHE university ranking of 2008 to evaluate the teaching efficiency of business administration departments of German universities. Since 1998, the CHE makes an annual ranking of German universities. The primary objective of the ranking is to inform potential students about the study conditions in German universities and to provide more transparency of the performance of the universities. Besides information about degree programs, university departments, and their locations, the ranking considers the perspective of university teachers and students. Therefore, the CHE conducts surveys among these two groups.

The main characteristics of the ranking are that

- it is subject specific,
- it is multidimensional, which means it comprises different indicators,
- it uses different perspectives, which means that besides facts, it uses personal judgments of professors and students, and
- it does not evaluate the universities on a scale, it rather builds three ranking groups for different criteria.

We use the indicators determined by the CHE ranking as inputs and outputs. The study is restricted to indicators that are crucial for the students' choice of a university with regard to study conditions. Contrary to the CHE

Table 1. Inputs and Outputs for Performance Evaluation (Berghoff et al., 2008).

Input	
Students	Total number of students in the respective degree course
Outputs	
Professors	The number of professors per department. An index for the size of the department and therefore the variety of classes and research activities
Library	This indicator shows the amount of money (in thousand euro) spent per year for books and scientific journals (including electronic journals) for a subject
Support from teachers	Graduates rated the support from teaching staff, both in general and in relation to the subjects, on a scale of 6 (very good) to 1 (very bad)
Courses offered	The students assessed the study situation in total in their faculty on a scale of 6 (very good) to 1 (very bad). This indicator is an individual question, not an index composed of different individual questions
Study organization	Index composed of several individual opinions. On a scale of 6 (very good) to 1 (very bad), the students assessed amongst other things the completeness of the courses offered in respect of the study regulations, the access opportunities to compulsory events, and the co-ordination of the courses offered with the examination regulations
Linkage theory practice	Index made up of several single judgments. Judgment of the students of cooperative education courses on the preparation- and follow-up-courses for the vocational training phases, the organization of these phases, and the quality of the supervision
Overall study situation	The students assessed the study situation in total in their faculty on a scale of 6 (very good) to 1 (very bad). This indicator is an individual question, not an index made up of different individual questions

ranking which is limited to three ranking groups for different indicators, we use the DEA estimator to create a ranking scale for the universities. The inputs and outputs used in our study are as noted in Table 1. The input *students* and the outputs *professors* and *library* are what the CHE calls facts, whereas the other outputs are student opinions determined by surveys. The students' opinion of each university were aggregated to an indicator on the university level, consequently the data consist of average values.

The available data set from the CHE ranking included university departments with missing data. Hence, at a first step, we eliminated these DMUs from the data set. Next, the data set was corrected for outliers by plausibility considerations, for example, we omitted one university department as it has only one professor. Finally, 50 universities remained for evaluation. They are characterized by the input and outputs described above.

We used DEA model (Eq. (6)) with a radial, output-oriented efficiency measure and VRS. According to Fandel (2003), the VRS technology is appropriate for the university sector. The confidence intervals, the bias-corrected efficiencies, and the empirical distribution functions were estimated by the homogenous bootstrap procedure presented in Chapter 9 with 2,000 bootstrap iterations.

3.2. Empirical Results

We applied the DEA model as well as the DEA bootstrap procedure to the data of the 50 university departments. Our intention is to show the differences in the interpretation of the original DEA estimator and the bias-corrected estimator. The results of the application are given in Table 2.

Applying the original DEA model to the data, we found 11 efficient DMUs, which are 22% of the university departments, the average DEA efficiency score is 1.103. When corrected for bias, average efficiency drops to 1.239 (with an average bias at 0.136).

A graphical representation of the results of the DEA model and the bootstrap procedure are shown in Fig. 1. The DMUs are ordered by their bias-corrected efficiency estimator. The original DEA estimators are represented by the circles and the bias-corrected bootstrap estimators by the dots. The 95% confidence intervals are indicated by the lower and the upper dashed lines. The figure reveals that the bias in many cases is quite large. This results in differences between the DEA estimator and the bias-corrected efficiency and leads to wide confidence intervals. For example, Kaiserslautern (DMU_{29}), initially estimated as DEA efficient, has the highest bias (0.272), and the lower and upper boundaries of its confidence interval are 1.015 and 1.551. The widths of the confidence intervals vary considerably and have a range between 0.125 (DMU_{26}, Hohenheim) and 0.536 (DMU_{29}, Kaiserslautern).

In general, efficiency of the DMUs is declining when correcting for a bias. This is due to the fact that applying the original DEA model efficiency is overestimated. However, the biases determined by the DEA bootstrap procedure are quite different over the DMUs. Obviously, the efficiency ranking of the DMUs changes when applying the bootstrap correction to the DEA estimators. This is especially the case for DMUs that in the first model were considered as efficient, that is, that yield a DEA score of 1. Their bias is on average 0.217, which is by far more than the overall average bias of 0.136. Some university departments that initially were evaluated as

Table 2. DEA Estimators and Bootstrap Results.

		DEA Estimator	Bias	Bias-Corrected Estimator	Confidence Interval		FSD-Efficiency
		$\hat{\theta}_i$	$\widehat{bias}(\hat{\theta}_i)$	$\hat{\hat{\theta}}_i$	2.5%	97.5%	$\hat{z}_i$
1	RWTH Aachen	1.150	0.132	1.1651	1.165	1.372	1.157
2	Uni Augsburg	1.132	0.089	1.221	1.146	1.299	1.103
3	Uni Bamberg	1.174	0.108	1.282	1.186	1.362	1.158
4	Uni Bayreuth	1.077	0.103	1.180	1.091	1.258	1.065
5	HU Berlin	1.141	0.143	1.284	1.162	1.400	1.159
6	TU Berlin	1.292	0.142	1.434	1.310	1.546	1.294
7	Uni Bielefeld	1.158	0.135	1.293	1.166	1.407	1.166
8	Uni Bochum	1.000	0.172	1.172	1.014	1.351	1.000
9	Uni Bremen	1.000	0.247	1.247	1.024	1.462	1.067
10	TU Chemnitz	1.235	0.075	1.310	1.249	1.387	1.183
11	TU Clausthal	1.31	0.133	1.163	1.040	1.314	1.049
12	TU Dortmund	1.289	0.113	1.402	1.312	1.490	1.266
13	TU Dresden	1.000	0.208	1.208	1.028	1.411	1.035
14	Uni Düsseldorf	1.128	0.128	1.256	1.139	1.396	1.133
15	Uni Duisburg-Essen/ Duisb.	1.055	0.120	1.175	1.0	1.312	1.060
16	Uni Duisburg-Essen/ Essen	1.070	0.086	1.156	1.080	1.232	1.044
17	Uni Eichst.-Ing./ Ingolstadt	1.000	0.176	1.176	1.27	1.277	1.000
18	Uni Erl.-Nürnb./ Nürnberg	1.200	0.081	1.3621	1.212	1.362	1.157
19	Uni Frankfurt/M.	1.037	0.159	1.196	1.047	1.359	1.077
20	Europ. Uni Frankfurt/ O.	1.095	0.095	11.1	1.105	1.265	1.075
21	TU Bergakademie Freiberg	1.025	0.102	1.127	1.037	1.238	1.018
22	Uni Greifswald	1.060	0.084	1.144	1.073	1.220	1.033

Table 2. (*Continued*)

		DEA Estimator	Bias	Bias-Corrected Estimator	Confidence Interval		FSD-Efficiency
		$\hat{\theta}_i$	$\widehat{bias}\left(\hat{\theta}_i\right)$	$\hat{\hat{\theta}}_i$	2.5%	97.5%	$\hat{z}_i$
23	Uni Halle-Wittenberg	1.221	0.103	1.324	1.236	1.410	1.196
24	Uni Hamburg	1.085	0.164	1.249	1.106	1.395	1.127
25	Uni Hannover	1.142	0.118	1.260	1.160	1.383	1.138
26	Uni Hohenheim	1.062	0.074	1.136	1.079	1.204	1.000
27	TU Ilmenau	1.063	0.140	1.203	1.082	1.330	1.086
28	Uni Jena	1.077	0.086	1.163	1.090	1.232	1.050
29	TU Kaiserslautern	1.000	0.272	1.272	1.015	1.551	1.000
30	Uni Kassel	1.000	0.226	1.226	1.029	1.394	1.000
31	Uni Kiel	1.013	0.095	1.108	1.032	1.192	1.000
32	Uni Köln	1.000	0.270	0.270	1.020	1.541	1.084
33	Uni Leipzig	1.046	0.158	1.204	1.062	1.332	1.085
34	Uni Magdeburg	1.163	0.102	1.265	1.180	1.342	1.142
35	Uni Mannheim	1.000	0.214	1.214	1.014	1.369	1.000
36	Uni Marburg	1.150	0.104	1.254	1.162	1.331	1.133
37	LMU München	1.000	0.166	1.166	1.014	1.271	1.000
38	TU München	1.014	0.161	1.175	1.034	1.313	1.059
39	Uni Münster	1.050	0.091	1.141	1.070	1.223	1.030
40	Uni Paderborn	1.235	0.103	1.338	1.255	1.419	1.208
41	Uni Passau	1.000	0.228	1.228	1.015	1.431	1.000
42	Uni Potsdam	1.278	0.101	1.379	1.288	1.467	1.245
43	Uni Regensburg	1.066	0.111	1.177	1.085	1.280	1.063
44	Uni Rostock	1.184	0.120	1.304	1.203	1.403	1.178
45	Uni Saarbrücken	1.135	0.094	1.229	1.1350	1.309	1.110
46	Uni Trier	1.239	0.119	1.358	1.255	1.449	1.226
47	Uni Tübingen	1.167	0.096	1.262	1.178	1.350	1.140
48	Uni Ulm	1.000	0.205	1.205	1.019	1.372	1.031
49	Uni Würzburg	1.203	0.104	1.308	1.223	1.405	1.181
50	Uni Wuppertal	1.221	0.111	1.332	1.241	1.431	1.203

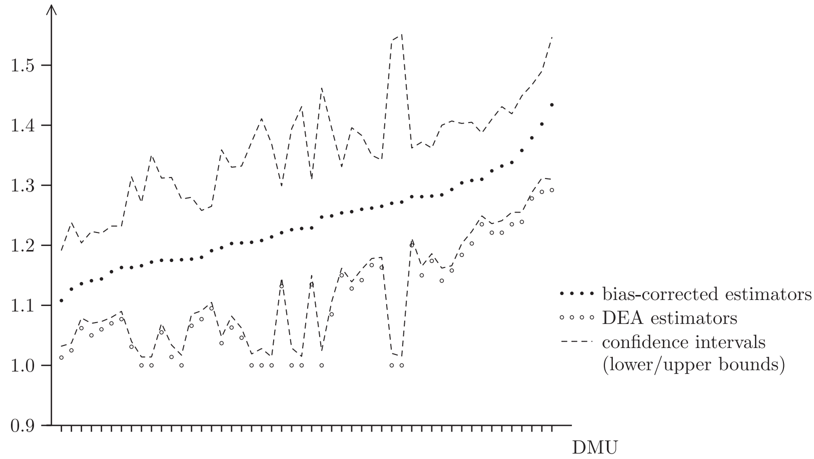


Fig. 1. DEA Estimators, Bias-Corrected Estimators and Confidence Intervals.

inefficient now are ranked higher than university departments that seemed to be efficient. For example, the university that is most efficient in its teaching performance is Kiel (DMU_{31}) according to its DEA bootstrap estimator (1.108), but Kiel was not efficient when applying the DEA estimator (1.013). Fig. 2 depicts a direct comparison of the original DEA estimator and the bias-corrected estimator.

The analysis of FSD supports the recorded findings. The department of Kiel (DMU_{31}) ranks amongst the best. In total, we find nine FSD-efficient universities. These departments are not stochastically dominated because a university with a higher probability for arbitrary values of efficiency estimator does not exist. Fig. 3 illustrates the empirical distributions of selected departments. It is obvious, that besides Kiel (DMU_{31}), Bochum (DMU_8) belongs to the reference set of FSD-efficient DMUs, for example. A closer examination reveals that Kiel is used as reference DMU for almost all FSD-inefficient departments. Consequently, inefficient universities should analyze Kiel's performance from students' perspective in detail. Moreover, Fig. 3 visualizes differences of inefficient departments. In contrast to Düsseldorf (DMU_{14}), the department of TU Berlin (DMU_6) is further away from the efficient reference and that is why TU Berlin is characterized by greater FSD-value $\hat{z}$ than Uni Düsseldorf. Finally, Spearman's rank correlation coefficient between corrected estimator and FSD value is 0.890, Spearman's rho 0.847.

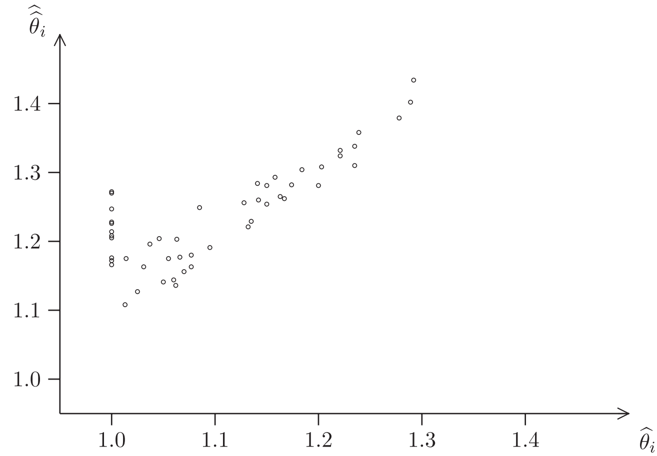


Fig. 2. DEA Estimators and Bias-Corrected Estimators.

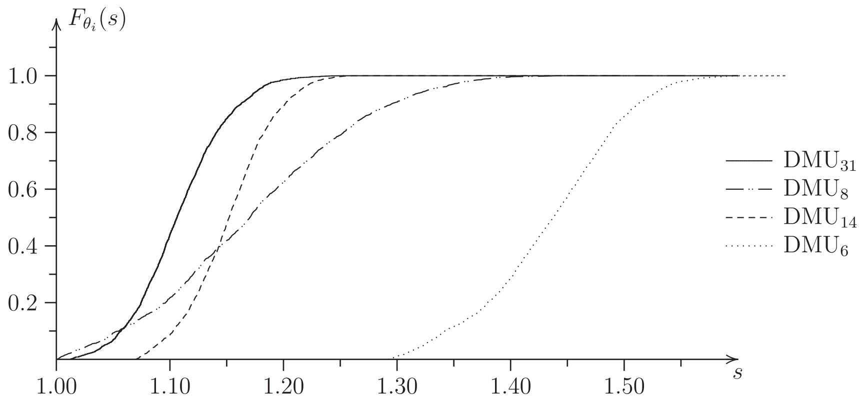


Fig. 3. Selected Empirical Distributions.

4. CONCLUSION AND FURTHER RESEARCH

We have looked at the teaching efficiency of German business administration departments. The evaluation was done from students' point of view as we intend to find out which universities offer good study conditions for the students. With the original DEA model, it was possible to identify efficient university departments and rank universities accordingly. But this ranking may not be reliable due to sampling variations. Thus, we suggested to employ a DEA bootstrap procedure to achieve higher reliability of the performance evaluation. The ranking of the university departments changed considerably when the efficiencies were corrected for bias. Departments that seemed to be efficient applying the original DEA model actually had the highest biases and were ranked lower than departments that were evaluated as inefficient. We used the concept of FSD to compare the estimated empirical distributions of the universities and to detect reference universities that may serve as benchmarks for inefficient universities.

This study focused on the students' perspective of teaching performance. For further applications, it would be interesting to reveal the differences when evaluating teaching performance from the other stakeholders' perspectives. The study also could be extended to other fields of interest for potential students, such as research performance or information on the study location. Consideration of additional universities – for example, by integration of Swiss and Austrian data – could lead to more reliable efficiency estimates and rankings as the reliability of the DEA bootstrap results increases by the number of considered DMUs.

ACKNOWLEDGMENT

The authors thank the *Centrum für Hochschuldidaktik* for the data.

REFERENCES

Al-Fahham, R. A. (2008). *Effizienz und Produktivität in deutschen Universitäten: Statische, dynamische und stochastisch basierte Anwendungen der Data Envelopment Analysis, dissertation*. Berlin: de-Verlag im Internet GmbH.

Berghoff, S., Federkeil, G., Giebisch, P., Hachmeister, C.-D., Hennings, M., Müller-Böling, D., & Roessler, I. (2008). *CHE Hochschulrankinge: Vorgehensweisen und Indikatoren-2008*. Working Paper no. 106. Centrum für Hochschulentwicklung, Gütersloh.

Charnes, A., Cooper, W., & Rhodes, E. (1978). Measuring efficiency of decision making units. *European Journal of Operational Research*, *2*(6), 429–444.

Cooper, W. W., Seiford, L. M., & Tone, K. (2007). *Data envelopment analysis* (2nd ed.). New York, NY: Springer.

Daraio, C., & Simar, L. (2007). *Advanced robust and nonparametric methods in efficiency analysis: Methodology and applications*. New York: Springer.

Dyckhoff, H., Rassenhövel, S., Gilles, R., & Schmitz, C. (2005). Beurteilung der Forschungsleistung und das CHE-Forschungsranking betriebswirtschaftlicher Fachbereiche. *Wirtschaftswissenschaftliches Studium*, *34*, 62–69.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York, NY: Chapman and Hall.

Fandel, G. (2003). Zur Leistung nordrhein-westfälischer Universitäten: Gegenüberstellung einer Verteilungslösung und der Effizienzmaße einer Data Envelopment Analysis. In: U. Backes-Gellner & C. Schmidtke (Eds), *Hochschulökonomie-Analysen interner Steuerungsprobleme und gesamtwirtschaftlicher Effekte* (pp. 33–50). Berlin: Duncker & Humblot.

Färe, R., Grosskopf, S., & Lovell, C. A. K. (1985). *The measurement of efficiency of production*. Boston: Kluwer.

Gijbels, I., Mammen, E., Park, B. U., & Simar, L. (1999). On estimation of monotone and concave frontier functions. *Journal of the American Statistical Association*, *94*(445), 220–228.

Gilles, R. (2005). *Performance measurement mittels data envelopment analysis: Theoretisches Grundkonzept und universitäre Forschungsperformance als Anwendungsfall*. Lohmar-Köln: Josef Eul Verlag.

Gutierrez, M. (2005). *Effizienzmessung in Hochschulen: Evaluation von Forschungs- und Lehreinheiten mit der Data Envelopment Analysis*. Wiesbaden: DUV.

Hanoch, G., & Levy, H. (1969). The efficiency analysis of choices involving risk. *Review of Economic Studies*, *36*, 335–346.

Kuosmanen, T. (2004). Efficient diversification according to stochastic dominance criteria. *Management Science*, *50*, 1390–1406.

Marx, J. (2003). *Mean Risk-Effizienz versus Stochastische Effizienz*. Frankfurt am: Lang.

Rhodes, E. L. (1978). *Data envelopment analysis and approaches for measuring the efficiency of decision making units with an application to program follow through in U.S. education*. Ph.D. dissertation, School of Urban and Public Affairs, Carnegie-Mellon University.

Sarrico, C. S., Hogan, S. M., Dyson, R. G., & Athanassopoulos, A. D. (1997). Data envelopment analysis and university selection. *Journal of the Operational Research Society*, *48*, 1163–1177.

Simar, L., & Wilson, P. W. (1998). Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Management Science*, *44*(1), 49–61.

Simar, L., & Wilson, P. W. (2000a). Statistical inference in nonparametric frontier models: The state of the art. *Journal of Productivity Analysis*, *13*, 49–781.

Simar, L., & Wilson, P. W. (2000b). A general methodology for bootstrapping in non-parametric frontier models. *Journal of Applied Statistics*, *27*(6), 779–802.

Warning, S. (2004). Performance differences in German higher education: Empirical analysis of strategic groups. *Review of Industrial Organization*, *24*, 393–408.

Wendell, R. E., & Lee, D. N. (1977). Efficiency in multiple objective optimization problems. *Mathematical Programming*, *12*, 406–414.

Worthington, A. C. (2001). An empirical survey of frontier efficiency measurement techniques in education. *Education Economics*, *9*(3), 245–268.

ASSESSMENT OF IMPLICATION OF COMPETITIVENESS ON HUMAN DEVELOPMENT OF COUNTRIES THROUGH DATA ENVELOPMENT ANALYSIS AND CLUSTER ANALYSIS

Füsun Ülengin, Özgür Kabak,
Şule Önsel and Emel Aktaş

ABSTRACT

Globalization speeds up competition among nations in various sectors. In terms of multinational and transnational phenomena, countries are seen as inescapable from competition, thus the linking of the term global with "competitiveness." The research described here explores the relationship between the competitiveness of a country and its implications for human development. For this purpose, using data envelopment analysis (DEA) and cluster analysis, 44 selected countries were evaluated. An output-oriented super-efficiency model where global competitiveness indicators are taken as input variables with human development indicators as output variables is utilized. Then cluster analysis depending on the competitiveness

Financial Modeling Applications and Data Envelopment Applications
Applications of Management Science, Volume 13, 199–226

ISSN: 0276-8976/doi:10.1108/S0276-8976(2009)0000013013

and human development indicators is conducted by using self-organizing maps to specify the development levels of the countries. Both analyses are repeated for years between 2005 and 2007. Finally, the relationship between the super efficiency scores and the development levels is analyzed.

1. INTRODUCTION

The near future is expected to bring important changes to the shape of the world economy and to the landscape of major industries. In seeking to explain patterns of international competition, several researchers such as Kogut (1988), Porter (1990), and Dunning (1990, 1993) have emphasized the importance of the characteristics of the home country in determining the competitive position of national firms in international markets. A nation's competitiveness has been defined by Artto (1987) as "the degree to which a nation can, under free and fair market conditions, produce goods and services that meet the test of international markets while simultaneously expanding the real incomes of its citizens."

If directed toward the needs of people, the competitiveness of a nation can bring advances for all humankind, but markets can go too far and squeeze the nonmarket activities that are so vital for human development: fiscal squeezes constrain the provision of social services; time squeezes reduce the supply and quality of labor; and incentive squeezes harm the environment. Globalization increases human insecurity as the spread of global crime, disease, and financial volatility outpaces actions to tackle them. Costantini and Monni (2007) state that human development has to be the first objective of international development policies, whereas an increase in human well-being is necessary to provide a sustainability path. Furthermore, a human development index (HDI) can be considered as a first and important step toward incorporating broad concepts of sustainability into measures of development (Sagar & Najam, 1998). In the Human Development Report of the United Nations Development Program (UNDP) (http://hdr.undp.org/hd/), the importance of a "human development" concept is stated thus: "The objective of development is to create an enabling environment for people to enjoy long, healthy and creative lives."

This study proposes a methodology to objectively analyze the relationship between the competitiveness level of a country and its capability to enhance human development. For this purpose, initially data envelopment analysis (DEA) is used. Then cluster analysis is conducted to make more meaningful interpretation of the DEA results.

Section 2 describes the state of the art in measuring the performance of nations in terms of both their competitiveness and their human development. Section 3 summarizes the basic stages used in the proposed methodology. In Section 4, 44 selected countries are evaluated on the basis of DEA analysis, where their efficiency in converting competitiveness into human development is investigated. These 44 countries are selected to represent 90% of world's GDP and population. In Section 5, a cluster analysis is conducted to classify the countries according to their World Economic Forum (WEF) and HDI scores. The relationship between the development and the efficiency of countries is also analyzed in this section. Finally, conclusions and further suggestions are given.

2. MEASURING THE PERFORMANCE OF NATIONS

Very few studies have attempted a comprehensive comparison of the performance of countries (Zanakis & Becerra-Fernandez, 2005). In an earlier study presented by Golany and Thore (1997), 72 countries were ranked by their economic and social performance. The inputs used to assess the performance of each country were domestic investments and government expenditures on both economic and social programs. The performance of countries across the world has been compared using various indices such as a global competitiveness index (Sala-i-Martin & Artadi, 2004). Several attributes are normally considered when developing those indices. The schemes used can be termed fixed-weight schemes, as they combine performance in terms of various attributes using pre-fixed weights, which may be chosen subjectively. Despotis (2005b), on the contrary, used a DEA-like index to estimate an ideal value of the composite HDI for each country. A goal programming model is also used in this study to obtain global estimates of human development. The new measure of human development is stated to be comparable and highly correlated with HDI. Lau and Lam (2002) propose a model that measures the economic freedom ranking of 161 countries. Their model consists of two stages. First, it calculates the set of weights for each country using DEA. Subsequently, conflicts among the criteria weights are resolved through a minimum disagreement decision model. Cherchye et al. (2008) proposed a DEA-based approach to create composite indicators, which are used for benchmarking countries' performances. The approach is illustrated using the Technology Achievement Index, which together with the HDI was developed by United Nations and included in the 2001 Human Development Report.

2.1. Evaluation of Competitiveness

Each year, some organizations such as the WEF and the Institute for Management Development (IMD) (http://www.imd.ch), publish rankings of national competitiveness among countries. These rankings serve as benchmarks for national policymakers and interested parties to judge the relative success of their country in achieving the competitiveness criteria represented by the corresponding competitiveness index. The IMD jointly with the WEF has produced comparisons of nations' competitiveness through the annual publication of the World Competitiveness Yearbook since 1989.

The WEF uses three competitiveness indices to analyze the competitiveness level from macroeconomic and microeconomic perspectives. The growth competitiveness index (GCI), developed by McArthur and Sachs (2001) and Blanke and Lopez-Claros (2004), makes an evaluation based on critical and, mostly, macroeconomic environmental factors that influence sustained economic growth over the medium to long term. Porter's business competitiveness index (BCI) (Porter, 1990), complementary to the GCI, investigates the company-specific factors that lead to improved efficiency and productivity indicators from a microeconomic perspective. Recently, a new index, the global competitiveness index (WEF, 2005), was designed with the goal of unifying the GCI and BCI and will eventually replace them in the global competitiveness report.

This new index is based on three principles: (1) The determinants of competitiveness are complex, competitiveness being composed of 12 pillars, each pillar having a different weight for each stage of development. (2) Economic development is a dynamic process of successive improvement, that is, it evolves in stages. In the most basic stage, called the *factor-driven stage*, firms compete on price and take advantage of cheap labor and/or unprocessed natural resources. In the second stage, called the *efficiency-driven stage*, efficient production becomes the main source of competitiveness. Finally, in the *innovation-driven* stage, successful economies can no longer compete on price or even quality and have to produce innovative products and practices using the most advanced methods of production and organization. (3) As economies develop, they move from one stage to the next in a smooth fashion (WEF, 2007, 2005).

The concept of "national competitiveness" has been criticized in recent years. According to some research, defending national competitive interests is often a façade for asking for privileges for particular groups, or seeking to prop up uneconomic activities (Lall, 2001). The WEF is clearly concerned with dynamic comparative advantage and emphasizes that the ability to sustain income and growth depends, in a globalized world, on each country's ability to innovate or import and use technologies created elsewhere. The WEF indices assign uniformly higher values to freer trade, stronger intellectual property protection, and more liberal capital accounts across countries. However, the WEF does not analyze whether a highly competitive country is also one that uses this power for the sake of human development (Lall, 2001). Thus, if the competitiveness of a country is properly managed, enhanced human welfare is the expected consequence. In fact, human development should be the ultimate objective of human activity and should be aimed at healthier, longer, and fuller lives (Ranis, Stewart, & Ramirez, 2000). However, there is a bi-directional link between human development and economic growth. Economic growth is seen as the major instrument for advancing human development. Achievements in human development can in turn also make a critical contribution to economic growth. In fact, human development is the ultimate goal of human activity (Anand & Sen, 2000).

2.2. Evaluation of Human Development

The original definition of human development was given in the UNDP Human Development Report (HDR) (UNDP, 1990, p. iii) as follows:

> Human development is a process of enlarging people's choices. In principle, these choices can be infinite and change over time. But at all levels of development, the three essential ones are for people to lead a long and healthy life, to acquire knowledge and to have access to resources needed for a decent standard of living. If these essential choices are not available, many other opportunities remain inaccessible.

The UNDP, through its Global HDRs, has defined development as a process of enlarging people's choices, as well as raising the level of well-being. In principle, these choices can be infinite and can vary over time and space. Among these, the HDRs identify the choices to lead a long and healthy life, to acquire knowledge and be educated, and to have access to the resources needed for a decent standard of living as the three most critical and socially valuable issues. Thus, the aforementioned report focuses on indicators of longevity, literacy, and per capita income (http://hdr.undp.org/hd/).

The UNDP claims that the HDI is superior to per-capita GDP for measuring social well-being because per-capita GDP measures only income, whereas the HDI is also weighted for longevity and education. Moreover, per-capita GDP only reflects average income, whereas the HDI is influenced by the type of goods that constitute the GDP. The HDI measures the average achievements in three basic dimensions of human development, namely, a long and healthy life, as measured by life expectancy at birth; knowledge, as measured by the adult literacy rate (with two-thirds weight) and a combined primary, secondary, and tertiary gross enrolment ratio (with one-third weight); and a decent standard of living, as measured by per-capita GDP.

Before the HDI itself is calculated, an index needs to be created for each of these dimensions. The details of the calculations are given in (www.un.org.my/uploads/files/HDI_Technical_note.pdf). HDR is one of the major contributions that oriented the debate on the measurement of development beyond the traditional economic perspective toward a broader scheme that incorporates different aspects of life into measure of human development (Despotis, 2005b).

2.3. Relationship between Competitiveness and Human Development

To the best of our knowledge, the relationship between competitiveness and human development has not been analyzed in depth in the literature. Ranis et al. (2000) investigated the connection between economic growth and human development, indicating a strong connection is expected. On the one hand, economic growth provides the resources to permit sustained improvements in human development. Besides, improvement in the quality of the labor force is an important contributor to economic growth. However, they acknowledged that while this two-way relationship may now be widely accepted, the specific factors linking the two elements have not been systematically explored. Davies and Quinlivan (2006) performed panel analysis on the impact of trade on human development. They state that the standard argument for a positive relationship between trade and human development is that more trade results in a greater standard of living, which, in turn, results in more education, better health care, better social services, etc. The standard argument rests on the premise that the influence of trade on income is direct, whereas the influence of trade on nonincome measures is indirect, being transmitted through income. Trade results not merely in an increase in the quantity of goods consumed, but in an increase in various goods consumed. In the case of a developing nation, new types of goods will include medicines, health-related equipment, and medical training – all of which improve the health, nutrition, and longevity of a country's people. Improvements in trade will result in some immediate economic gains. The immediate economic gains will, in turn, result in future increases in literacy and health as people's standards of living rise and the opportunities for returning to education increase. Davies and Quinlivan (2006) find that increases in trade are also positively associated with future increases in social welfare.

3. FRAMEWORK OF THE PROPOSED METHODOLOGY

Our proposed methodology, comprising two stages, aims to determine how efficient countries are in using their competitiveness for the improvement in human development. In the first stage, DEA analysis is conducted. The aim of this stage is to measure the efficiency of countries in using their competitiveness for the welfare of their citizens. For this purpose, the scores for the three main dimensions of the WEF's competitiveness index, namely, basic requirements (BASREQ), efficiency enhancers (EFFENH), and innovation and sophistication factors (INSOPF), are used as the inputs. On the contrary, to represent the human development of the countries, the criteria used in the HDI are taken as the output of the DEA. These criteria are life expectancy at birth (LIFEXP), combined gross enrolment ratio for primary, secondary, and tertiary schools (ENROLL), and GDP per capita (GDPCAP), representing the three main dimensions of human development: health, education, and economy. Input and output data were gathered from WEF (2004, 2005, 2006) and UNDP (2006, 2007, 2008), respectively.

In the second stage, a cluster analysis is conducted utilizing artificial neural networks (ANN), namely self-organized maps (SOM). The classification is based on aforementioned WEF and HDI indicator scores. The development level of the countries is specified by the cluster analysis.

Finally, efficiency scores and the development levels of the countries for the data of different years (i.e., yearly data from 2005 to 2007) are analyzed to identify the evolution of the countries from the competitiveness and human development perspectives. Fig. 1 shows the detailed framework of the proposed methodology.

3.1. Data Envelopment Analysis

The basic techniques that have been used for measuring efficiency can be categorized into econometric and mathematical programming approaches.

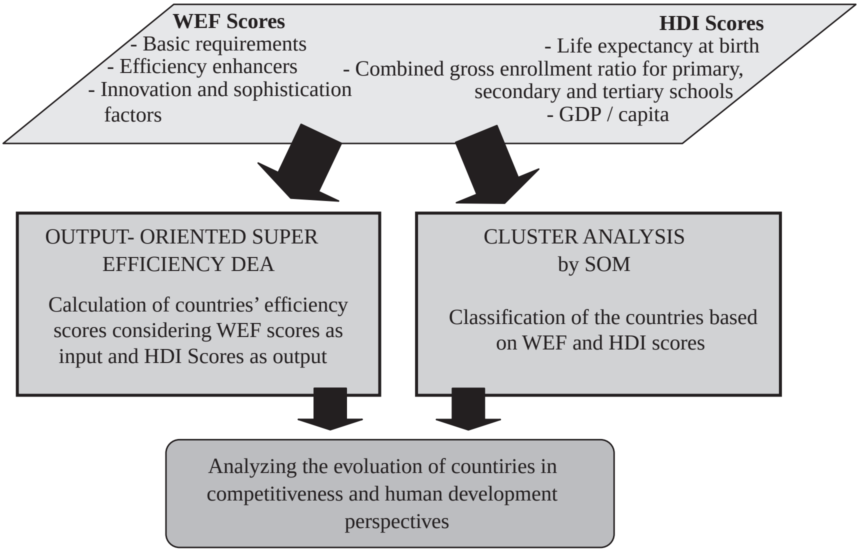


Fig. 1. Framework of Proposed Methodology.

The former includes regression-based techniques and is stochastic; the stochastic frontier analysis being one of most widely applied econometric techniques. The mathematical programming approaches used in the production frontier analysis include a wide range of nonparametric techniques that are largely nonstochastic such as DEA and goal programming (Fried, Lovell, & Schmidt, 1993).

DEA is a method of measuring the relative efficiency for a group of operating units where the relative values of the variables are unknown. It accommodates multiple inputs and outputs and can also include exogenously fixed environmental variables. DEA utilizes the fundamental concept of a production function, and since it uses linear programming (LP), it is a nonparametric technique that does not require assumptions about the statistical properties of the variables (Retzlaff-Roberts, Chang, & Rubin, 2004).

DEA is one of the most well-known technique for measuring the relative efficiency of decision-making units (DMUs) on the basis of multiple inputs and outputs. The efficiency of a unit is defined as the weighted sum of its outputs divided by a weighted sum of its inputs, and it is measured on a bounded ratio scale. The weights for inputs and outputs are estimated by a linear program in the best advantage for each unit so as to maximize its relative efficiency. Basically, DEA provides a categorical classification of the units into efficient and inefficient ones by assuming either constant or variable returns to scale for the inputs and outputs. DEA is currently used in various fields to measure the performance of diverse entities, considered as DMUs (Despotis, 2005a, 2005b). Gattoufi, Oral, Kumar, and Reisman (2004) can be seen for content analysis of DEA literature and its comparison with that of other OR/MS fields.

In DEA, it is a normal practice to decide the relative importance of competing explanatory factors before the analysis. The inputs and the outputs are entered into the DEA optimization algorithm, but there is no built-in test of their appropriateness. DEA does not require the specification of a functional form to be fitted. If the true functional form is unknown, this feature of DEA could be advantageous, since it avoids the danger of fitting the wrong functional form. If used carefully with large samples, DEA is good at identifying possible reasons for apparently poor performance, highlighted by crude indicators such as performance ratios, and is good at providing a checklist of questions for management (Cubbin & Tzanidakis, 1998).

Although the basic idea of DEA is well founded and clearly grounded in economic theory, the practicality of choosing appropriate inputs and outputs in the context of performance measurement is complex. Management opinion may not necessarily be given in the form of explicit identification of input/output factors; rather, it is often expressed in the more global sense of DMUs being efficient or inefficient. In many circumstances, this form of expression of expertise can be a valuable input to a performance measurement exercise.

Charnes–Cooper–Rhodes (CCR) model (Charnes, Cooper, & Rhodes, 1978) and Banker–Charnes–Cooper (BCC) models (Banker, Charnes, & Cooper, 1984) are two basic DEA models, where the former exhibit constant returns to scale (CRS) and the latter exhibit variable returns to scale (VRS). VRS model has been developed specifically to accommodate scale effects in analysis. However, when there are no inherent scale effects, small and large units will tend to be over-rated in the efficiency assessment because VRS model will always envelop the data more closely than the CRS model (Dyson et al., 2001).

The efficiency score of a DMU is measured by means of a combination of DEA-efficient DMUs, which form part of the segments on the efficiency frontier. The efficient DMUs are not comparable among themselves in the CCR model or other DEA models. To avoid incomparability of DMUs, a super-efficiency DEA model (Li, Jahanshahloo, & Khodabakhshi, 2007) was used in this study.

3.1.1. Theoretical Framework of Super-Efficiency Data Envelopment Analysis
DEA is a data-oriented method for measuring and benchmarking the relative efficiency of peer DMUs (Charnes et al., 1978; Cooper, Seiford, & Tone, 2000). We assume that there are n homogeneous DMUs such that all the DMUs use m inputs x_{ij} ($i = 1, \ldots, m$) to produce s outputs y_{rj} ($r = 1, \ldots, s$). It is also assumed that $X_j = (x_{ij}) \in \Re^{s \times n}$ is nonnegative. An output-oriented CCR DEA model can thus be written as follows (Cooper et al., 2000):

Maximize η_0, subject to

$$\begin{aligned} x_{i0} - \sum_{j=1}^{n} \mu_j x_{ij} \geq 0 \quad & i = 1, \ldots, m \\ \eta_0 y_{r0} - \sum_{j=1}^{n} \mu_j y_{rj} \leq 0 \quad & r = 1, \ldots, s \\ \mu_j \geq 0 \quad & j = 1, \ldots, n \end{aligned} \tag{1}$$

where $1/\eta_0$ gives the efficiency value for DMU_0. To find all efficiency values of the DMUs, the LP model given above should be solved for each DMU.

In standard DEA models, a DMU is said to be efficient if its performance relative to other DMUs cannot be improved. In the absence of price data or preferential weightings of inputs and outputs, all efficient DMUs have equal scores of 1.0, and rank equally in terms of performance. Inefficient DMUs have scores less than 1.0 with an input orientation, and greater than 1.0 with an output orientation (Lovell & Rouse, 2003). In some problems, the number of efficient DMUs (i.e., with efficiency = 1.0) may be very high, resulting in incomparability of many DMUs. In such cases, a super-efficiency DEA model, a model in which the DMU under evaluation is excluded from the reference set of the original DEA model (Andersen & Petersen, 1993; Seiford & Zhu, 1998), is used. On the basis of the output-oriented CCR model defined earlier, an output-oriented super-efficiency model can be defined as follows:

Maximize η_0, subject to

$$\begin{aligned} x_{i0} - \sum_{\substack{j=1 \\ j \neq 0}}^{n} \mu_j x_{ij} \geq 0 \quad & i = 1, \ldots, m \\ \eta_0 y_{r0} - \sum_{\substack{j=1 \\ j \neq 0}}^{n} \mu_j y_{rj} \leq 0 \quad & r = 1, \ldots, s \\ \mu_j \geq 0 \quad & j = 1, \ldots, n \end{aligned} \tag{2}$$

Here $1/\eta_0$ again gives the efficiency value for DMU_0. For the inefficient DMUs (i.e., those with efficiency <1), scores in the previous model are the same with the ones in this super-efficiency model. However, the scores of the efficient DMUs in the previous model (i.e., those with efficiency $=1$) are greater than 1 in the super-efficiency model, which allow a comparison of efficient DMUs.

3.2. Cluster Analysis

Cluster analysis involves grouping similar objects into mutually exclusive subsets referred as clusters (Hair, Anderson, & Black, 1995). Also called as segmentation analysis or unsupervised classification, cluster analysis is a method of creating groups of objects, or clusters in such a way that objects in one cluster are very similar and objects in different clusters are quite different (Gan, Ma, & Wu, 2007). Cluster analysis is often confused with classification, in which objects are assigned to predefined classes. In cluster analysis, the classes are also to be defined.

The cluster definition problem is NP-complete. As a result, an optimum does not exist. A number of heuristic methods including agglomerative techniques, which are the mostly widely known and used, have been built for this purpose. All hierarchical agglomerative heuristics begin with n clusters where n is the number of observations. Then, the two most similar clusters are combined to form $n-1$ clusters. On the next iteration, $n-2$ clusters are formed with the same logic, and this process continues until one cluster remains. Only the rules used to merge clusters differ in various hierarchical agglomerative heuristics. The "Simple Linkage" approach merges the clusters by finding the minimum distance between one observation in one cluster and another observation in the second cluster. "Furthest Neighborhood," in contrast, takes the furthest distance between two observations, whereas "Average Linkage" takes the average distance of the observations belonging to each cluster and merges them with a minimum average distance between all pairs of observations in the respective clusters. In Ward's method, on the contrary, the distance is the ANOVA sum of squares between the two clusters summed over all variables (Onsel, Ulengin, & Ulengin, 2004).

Although all hierarchical methods successfully define clusters for compact and isolated data, they generally fail to accurately provide defined clusters for "messy" data. The major issue with all clustering techniques is how to select the number of clusters. Different clustering methods may lead to different clusters, and the differences are generally due to the inherent characteristics of the methodology used. In fact, there is no single methodology that can be recommended in selecting the most appropriate number of clusters and the most suitable clustering method. That is why cluster analysis is generally accepted to be more of an art than a science (Milligan, 1980).

3.2.1. Self-Organizing Maps

To improve the accuracy of the cluster analysis and to reduce the subjectivity that plays an important role in hierarchical clustering, the self-organizing map (SOM) Neural network is used as suggested by Mangiameli, Chen, & West (1996). SOM learns to detect groups of similar input vectors in such a way that neurons physically close together in the neuron layer respond to a similar input vector (Kohonen, 1987). They learn both the distribution and the topology of the input vectors they are trained on. They are unsupervised networks; that is, they have no output value in the training pattern to which training can be compared. In most other network models, all neurons adjust their weights in response to a training presentation while in SOM, that is not the case. In this kind of network, the neurons compete for the privilege of learning. SOM networks have two layers, the input layer of N variables and an output layer, which has one output for each category. During training, the patterns are presented to the network, then propagated to the output layer and evaluated. The network adjusts the weights to the output neurons in a neighborhood around the neuron. The neighborhood size is variable; it begins large and slowly reduces. SOM networks work by clustering patterns based on their distance from each other.

4. EVALUATION OF SELECTED COUNTRIES BASED ON DATA ENVELOPMENT ANALYSIS

In this study, to make a comparison between countries of similar scale, countries representing 90% of the world in terms of both world population and total world GDP are selected for evaluation. For this reason, among the countries evaluated by the WEF, only those having a population of over 25 million and/or a GDP level over US$200 billion were selected (the selection is made according to the data given in WEF, 2004). Of the selected countries, 17 constituted the largest economies in the world in terms of purchasing power parity (PPP). These were the Group of Seven (G7) industrialized countries (United States, Japan, Germany, United Kingdom, France, Italy, and Canada), as well as Spain, Australia, and Korea, and the seven largest emerging-market economies, referred to collectively as "E7" (China, India, Brazil, Russia, Indonesia, Mexico, and Turkey) (Hawksworth, 2006). The selected countries also include 20 Organisation for Economic Co-operation and Development (OECD) countries, 12 European countries that are members of the European Union, and four European countries not members of the European Union (Turkey, Russia, Norway, and Ukraine). There were also 10 countries from Africa, 9 countries from Asia, 6 from South America, 2 from North America, and Australia.

As mentioned in Section 3, the inputs of the DEA model were the BASREQ, EFFENH, and INSOPF; the outputs were LIFEXP, ENROLL, and GDPCAP. The application of DEA presents a range of issues relating to the homogeneity of the units under assessment, the input/output set used, the measurement of those variables, and the weights attributed to them in the analysis. Each of these issues can present practical difficulties in applying DEA. When constricting a DEA model, these assumptions should be considered. To avoid the pitfalls highlighted by Dyson et al. (2001), in this research, all the selected inputs are index measures while the outputs are volume measures. In order not to mix indices with volume measures for the outputs, the literacy rate indicator of the HDI was omitted. In fact ENROLL can be accepted as sufficient to measure the education level of a country. To mix indices, often associated with performance measures, with activity levels, which are volume measures is generally not suitable in DEA. This may be acceptable if all the inputs and outputs are of the same kind, as is proposed in this study (Dyson et al., 2001).

The original assumption on the measurement scales of the inputs and outputs is that they should conform to ratio scale. This in fact may be an unnecessarily strong assumption, but certainly an interval scale is an assumption of many DEA models (Dyson et al., 2001). In this study the inputs and one of the outputs, namely ENROLL, are measured in interval scale, whereas the other outputs (i.e., LIFEXP and GDPCAP) are measured in ratio scale.

The inputs, BASREQ, EFFENH, and INSOPF, are generated from 177 criteria, which are the hard data and survey data used in the WEF reports (WEF, 2004, 2005, 2006). As the survey data are in 1–7 Likert scale and the hard data are transformed to the same scale; the resulting indices are all in 1–7 scale. Therefore, they can be considered as index variables.

As far as the outputs are concerned, LIFEEXP data are based on the estimates gathered from World Population Prospects, the official source of United Nations population estimates and projections. ENROLL data are produced by the United Nations Educational, Scientific and Cultural Organization (UNESCO) Institute for Statistics based on enrolment data collected from national governments and population data from the United Nations Population Division's report. The gross enrolment ratios are calculated by dividing the number of students enrolled in all levels of schooling (excluding adult education) by the total population in the official age group corresponding to these levels. ENROLL can be over 100% due to the inclusion of over-aged and under-aged pupils/students because of early or late entrants, and grade repetition. That is why this variable can be considered as a volume measure. GDPCAP (PPP US$) data for the HDI are provided by the World Bank based on price data from the latest International Comparison Program surveys and GDP in local currency from national accounts data. PPPs for these countries are estimated directly by extrapolating from the latest benchmark results. For countries not included in the International Comparison Program surveys, estimates are derived through econometric regression. For countries not covered by the World Bank, PPP estimates provided by the Penn World Tables of the University of Pennsylvania are used (UNDP, 2007).

The data used are based on WEF and HDI and are accepted as reliable. In fact, the WEF data (i.e., the inputs) used in this study represent the best available estimates from various national authorities, international agencies, and private sources at the time the WEF Report (WEF, 2005) was prepared (July/August 2004). The WEF scores are generated from 181 indicators, 155 of which are the survey data. It is generally argued that the measurement of qualitative factors is highly subjective and using such data to characterize qualitative variables may result in an unfair DEA evaluation. However, the WEF survey data cover a large number of respondents, which reduces the effect of subjectivity on the measurement process. In fact, in the countries evaluated in this research the average number of respondents is 289.14.

The data of HDI (i.e., the outputs), on the contrary, are gathered from major statistical agencies. These are often specialized agencies of the United Nations working on issues such as health, World Health Organization (WHO); education, UNESCO Institute for Statistics; and labor market statistics, International Labor Organization (ILO) (http://hdr.undp.org/hdr2006/statistics/understanding). Despite some data availability problems, it is accepted that the HDI scores are internationally comparable (UNDP, 2007).

DEA is conducted using the data of several years. Firstly, the data supplied from WEF (2004) and UNDP (2006) are considered (i.e., data of year 2005). Secondly, WEF (2005) and UNDP (2007) data are taken into account (i.e., data of year 2006). And finally WEF (2006) and UNDP (2008) data are used in the DEA model (i.e., data of year 2007). The results are presented in Table 1.

It can be seen from Table 1 that Algeria, Argentina, Australia, Italy, Norway, United States, and Venezuela have been efficient countries in all the years considered. Belgium, Poland, and United Kingdom were efficient in 2005, and then became inefficient in 2006 and 2007. Brazil, Spain, and Ukraine have become efficient in 2007. Peru was only efficient in 2006.

5. CLUSTERING BASED ON WEF AND HDI INDICATORS

The aim of cluster analysis conducted in this study is to specify the general development stage of the countries and to analyze the relationship between their development stage and their efficiency. Then, the paths followed by the countries during their evolution is tried to be revealed.

Cluster analyses have been conducting using WEF indicators (i.e., BASREQ, EFFENH, and INSOPF) and HDI indicators (i.e., LIFEXP, ENROLL, and GDPCAP) for each year to see different groups that the countries may belong. For this purpose, SOM and the MATLAB software is used to determine the clusters that countries belong.

Since SOMs are unsupervised networks, the first decision has to be about the determination of the number of clusters that will group the countries according to their WEF and HDI scores. To this, an iterative approach is used. Initially, the WEF and HDI data are given as the input to SOM and through different topologies. These topologies account for different number of clusters. To determine the topology, the iteration starts with a high number of neurons (and hence clusters) and then the number of sample hits for the clusters (which are in fact the neurons of SOM) is checked out. If there are three or less sample hits, iteration proceeds with new SOMS that have less number of neurons (i.e., a smaller topology). This iteration is applied until all the clusters of the topology have at least four or more sample hits.

For all years, the first topology used is a 2*4 one indicating 8 neurons. On the basis of the assumption that, it is best if the data are fairly evenly distributed across the neurons, the process is conducted iteratively and the last decision given about the number of clusters that the countries should be

Table 1. Results of DEA Scores for Each Analysis Year.

	2005	2006	2007
Algeria	1.038	1.115	1.033
Argentina	1.118	1.025	1.077
Australia	1.026	1.079	1.067
Austria	0.891	0.917	0.928
Bangladesh	0.943	0.974	0.967
Belgium	1.013	0.938	0.936
Brazil	0.987	0.983	1.041
Canada	0.907	0.897	0.934
China	0.842	0.893	0.913
Colombia	0.995	0.940	0.955
Denmark	0.895	0.884	0.910
Egypt	0.853	0.916	0.934
Ethiopia	0.794	0.888	0.901
France	0.870	0.884	0.926
Germany	0.830	0.844	0.852
India	0.795	0.824	0.808
Indonesia	0.835	0.892	0.905
Italy	1.166	1.131	1.104
Japan	0.876	0.909	0.893
Kenya	0.723	0.798	0.829
Korea	0.851	0.870	0.880
Mexico	0.958	0.963	0.943
Morocco	0.846	0.979	0.925
Netherlands	0.905	0.926	0.896
Nigeria	0.721	0.719	0.761
Norway	1.205	1.224	1.158
Pakistan	0.949	0.993	0.934
Peru	0.994	1.067	0.982
Philippines	0.998	0.997	0.978
Poland	1.021	0.965	0.957
Russian Federation	0.962	0.980	0.981
South Africa	0.763	0.806	0.815
Spain	0.943	0.971	1.003
Sweden	0.936	0.880	0.855
Switzerland	0.850	0.894	0.895
Tanzania	0.755	0.734	0.825
Thailand	0.815	0.805	0.802
Turkey	0.889	0.989	0.944
Uganda	0.841	0.913	0.926
Ukraine	0.957	0.972	1.015
United Kingdom	1.077	0.902	0.887
United States	1.081	1.092	1.114
Venezuela	1.079	1.067	1.059
Vietnam	0.959	0.955	0.989

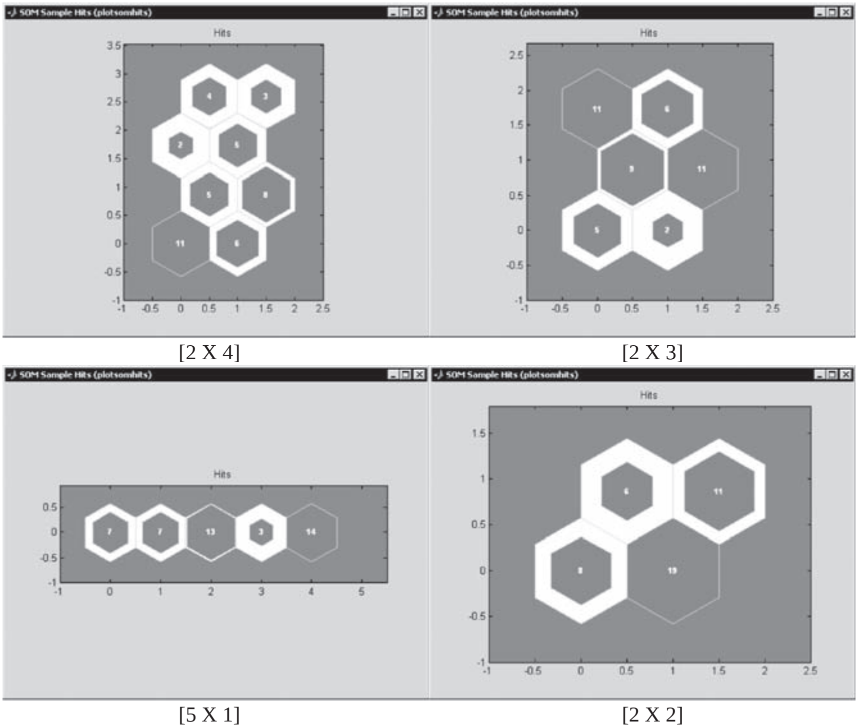


Fig. 2. Cluster Evaluation for 2005 Data.

grouped under is decided as 4 through a 2∗2 topology. It can be seen from Fig. 2 that there are two clusters with very few sample hits in the [2×4] neuron topology where there are eight clusters. So the iteration proceeds with [2×3] neuron topology where there are six clusters (since $8 - 2 = 6$ neurons are needed). It is found that there is only one cluster with only two sample hits in the [2×3] neuron topology, so the iteration proceeds with a [5×1] neuron topology (since $6 - 1 = 5$ neurons are needed). It is also found there is only one cluster with few sample hits in this topology so the iteration proceeds with a [2×2] neuron topology. There are four clusters in [2×2] neuron topology which are found to have enough sample hits and the iteration is ended.

The same iteration process is also conducted for 2006 and 2007 data, and as can be seen from Figs. 3 and 4 for both years the number of clusters that

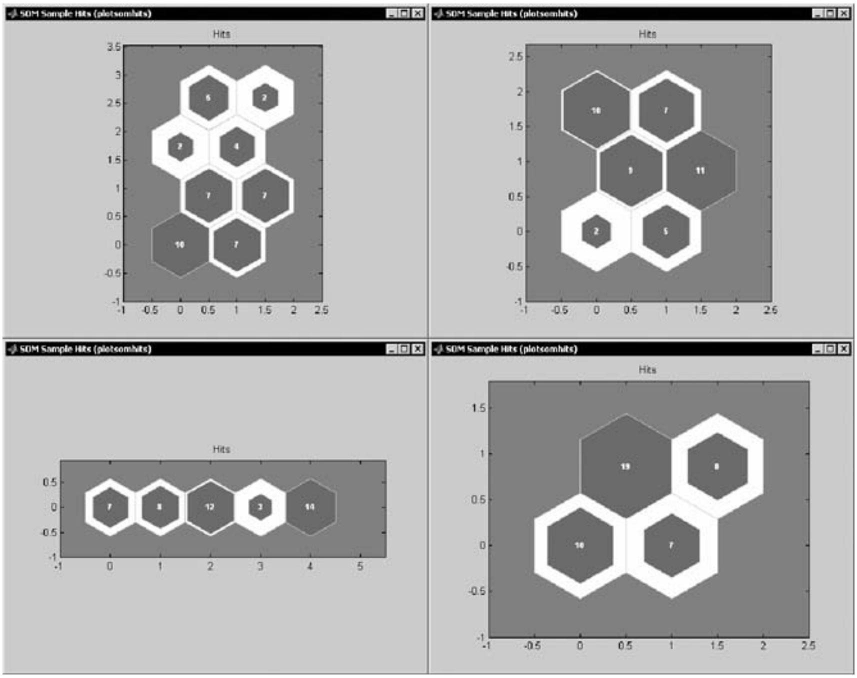


Fig. 3. Cluster Evaluation for 2006 Data.

the countries should be grouped under is decided as 4 through a [2×2] neuron topology.

As it can be seen from Fig. 4, the iteration proceeds from a [2×4] neuron topology to a [5×1] neuron topology since there are three neurons in [2×4] topology with very few sample hits. The appropriate number of clusters resulting from this initial stage is next used to carry out the clustering analysis to group the countries. Since we ought to categorize the countries into four classes, there are four outputs in the neural networks configuration. This leads to a 2×2 matrix of the weight vector. The topology function used is "HEXTOP," which means that the neurons are arranged in hexagonal topology at Kohonen layer, whereas the distance function is "MANDIST," which means that the used distance function is Manhattan distance (city block distance). Table 2 provides the clusters of the countries for all analysis years (efficient countries are shown in italic).

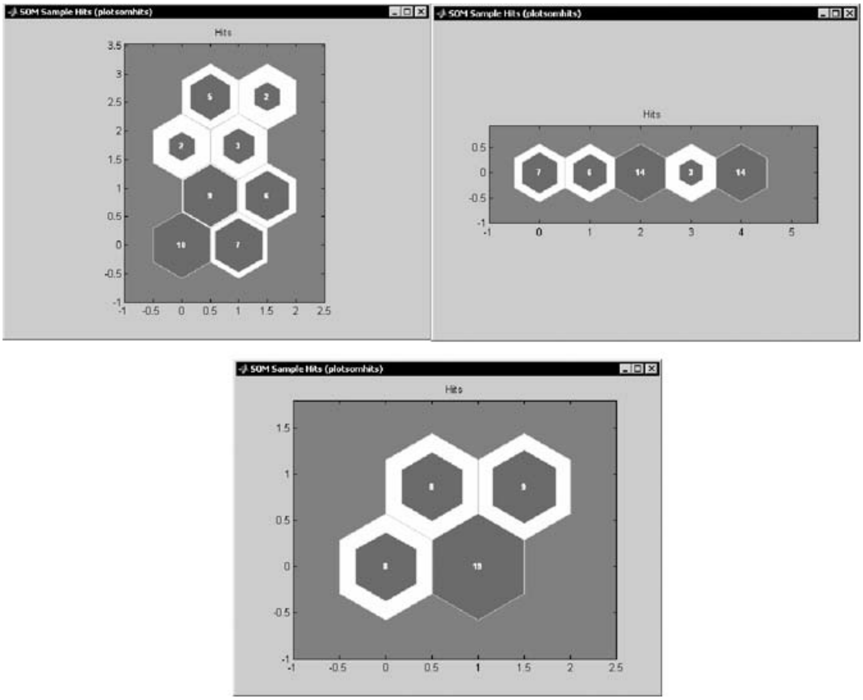


Fig. 4. Cluster Evaluation for 2007 Data.

5.1. Countries Cluster 4: Evolution from 2005 to 2007

When the results of cluster analysis are analyzed, it can be seen that according to 2005 results, the countries belonging to cluster 4, have the highest WEF and HDI indicator scores (Table 3). However, when those countries are analyzed according to their DEA score, it can be seen that Australia, Norway, United Kingdom, and Unites States are efficient (Tables 1 and 2) in transforming its competitiveness level into living conditions.

Average efficiency for this cluster is higher than all other clusters. However, when the path over the years followed by those countries is investigated, it can be seen that Australia, Canada, and Norway moved to third cluster. In addition to this United Kingdom has become inefficient. This shows that although United Kingdom has a high level of WEF and HDI scores, with the WEF score it has, it is expected that it should have a much higher HDI score.

Table 2. Clusters for Each Analysis Year.

2005				2006				2007			
1	2	3	4	1	2	3	4	1	2	3	4
Bangladesh	*Algeria*	Austria	*Australia*	Bangladesh	*Algeria*	*Australia*	Austria	Bangladesh	*Algeria*	*Australia*	Denmark
Ethiopia	*Argentina*	*Belgium*	Canada	Ethiopia	*Argentina*	Belgium	Denmark	Ethiopia	*Argentina*	Austria	Germany
Kenya	Brazil	France	Denmark	Kenya	Brazil	Canada	France	Kenya	*Brazil*	Belgium	Japan
Nigeria	China	*Italy*	Germany	Nigeria	China	*Italy*	Germany	Nigeria	China	Canada	Netherlands
Pakistan	Colombia	Korea	Japan	Pakistan	Colombia	Korea	Japan	Pakistan	Colombia	France	*Norway*
S. Africa	Egypt	Spain	Netherlands	S. Africa	Egypt	*Norway*	Netherlands	S. Africa	Egypt	*Italy*	Sweden
Tanzania	India		*Norway*	Tanzania	India	Spain	Sweden	Tanzania	India	Korea	Switzerland
Uganda	Indonesia		Sweden	Uganda	Indonesia		Switzerland	Uganda	Indonesia	*Spain*	UK
	Mexico		Switzerland		Mexico		UK		Mexico		*US*
	Morocco		*UK*		Morocco		*US*		Morocco		
	Peru		*US*		*Peru*				Peru		
	Philippines				Philippines				Philippines		
	Poland				Poland				Poland		
	Russian Fed.				Russian Fed.				Russian Fed.		
	Thailand				Thailand				Thailand		
	Turkey				Turkey				Turkey		
	Ukraine				Ukraine				*Ukraine*		
	Venezuela				*Venezuela*				*Venezuela*		
	Vietnam				Vietnam				Vietnam		

Table 3. Average Values for WEF, HDI, and Efficiency Scores (2005).

Clusters	WEF Scores			HDI Scores			**Average Efficiency**
	Average BASREQ	Average EFFENH	Average INSOPF	Average LIFEXP	Average ENROLL	Average GDPCAP	
1	3.88	2.92	2.86	50.71	54.13	2386.13	0.81
2	4.45	3.17	3.05	70.16	76.21	6238.95	0.94
3	5.26	3.92	3.91	79	94.83	25597.83	0.96
4	5.69	4.44	4.51	79.32	100.45	30595.36	0.96
Total	4,76	3,54	3,49	70,12	80,79	14267,39	**0.93**

In 2006, in cluster 4, the only efficient country is United States. All other countries (Austria, Denmark, France, Germany, Japan, Netherlands, Sweden, Switzerland, United Kingdom) do not show sufficient performance in transforming their competitiveness into better living conditions. In fact, Norway returns to cluster 4 in 2007, and for this year, Norway and the United States are efficient countries.

5.2. Countries Cluster 3: Evolution from 2005 to 2007

The general characteristic of cluster 3 countries is that they have second highest averages in terms WEF and HDI scores. Among those countries, Belgium and Italy are efficient in 2005. Canada and Australia belonged to cluster 4 in 2005; however, in 2006 and 2007, they were found to be cluster 3 countries. Similarly France and Austria moved from this cluster to cluster 4 in 2006, but they returned back in 2007 (Fig. 5).

5.3. Countries Cluster 2: Evolution from 2005 to 2007

The countries corresponding to cluster two show steady state in terms of cluster membership in the 2005–2007 period. These countries are third in terms of average values of WEF and HDI.

Algeria, Argentina, and Venezuela were efficient countries for all years. Although their competitiveness levels are lower when compared to clusters 3 and 4, they manage well to transform their competitiveness into better living conditions.

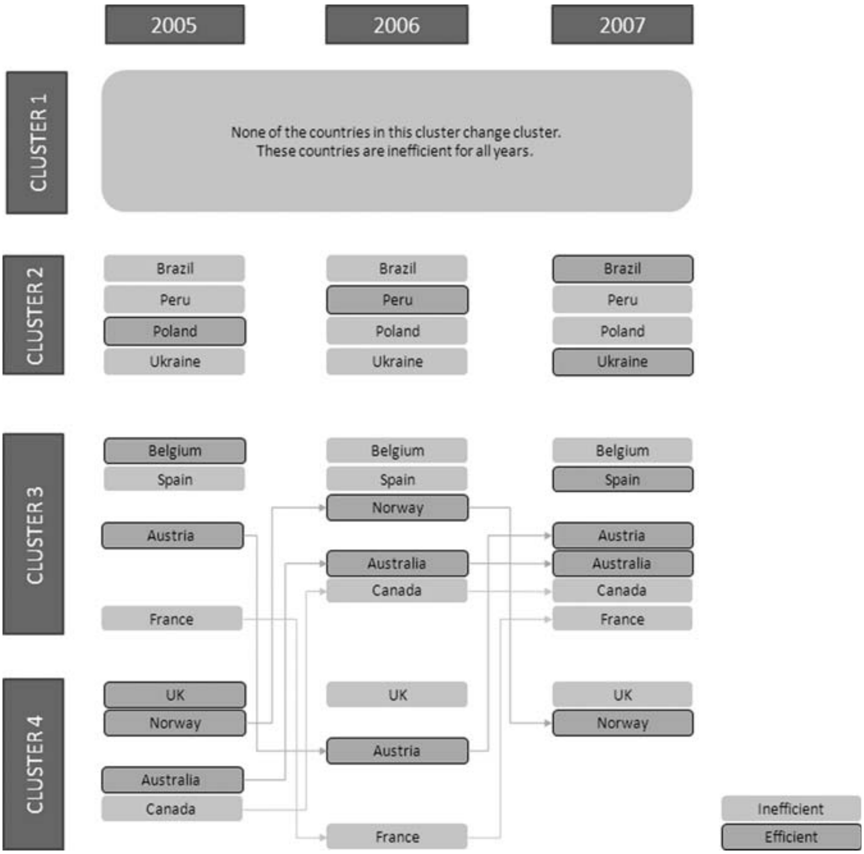


Fig. 5. The Countries Changing Cluster and Efficiency.

5.4. Countries Cluster 1: Evolution from 2005 to 2007

The countries in this cluster have consistently low performance in all years. They have low WEF and HDI scores, and they are all inefficient in transforming their low competitiveness level into better living conditions in all years.

Fig. 5 shows the changes in efficiency status as well as cluster membership of the countries over 2005–2007 period.

It can be seen from Tables 3–5 that clusters 3 and 4 are always above the general average, whereas clusters 2 and 1 are always below the general average for WEF and HDI scores.

Table 4. Average Values for WEF, HDI, and Efficiency Scores (2006).

Clusters	WEF Scores			HDI Scores			**Average Efficiency**
	Average BASREQ	Average EFFENH	Average INSOPF	Average LIFEXP	Average ENROLL	Average GDPCAP	
1	3.87	3.15	3.4	50.84	54.63	2561.13	0.85
2	4.47	3.69	3.64	70.39	75.63	6869.05	0.96
3	5.52	5.03	4.81	79.51	97.29	29267.14	1.02
4	5.74	5.37	5.59	79.27	92.5	31591.1	0.91
Total	4.82	4.19	4.23	70.3	79.09	15267.77	**0.94**

Table 5. Average Values for WEF, HDI, and Efficiency Scores (2007).

Clusters	WEF Scores			HDI Scores			**Average Efficiency**
	Average BASREQ	Average EFFENH	Average INSOPF	Average LIFEXP	Average ENROLL	Average GDPCAP	
1	3.71	3.28	3.48	53.7	55.66	2644.25	0.87
2	4.45	3.81	3.72	71.08	76.15	7381.37	0.96
3	5.48	5.01	4.89	79.79	97.54	29887.63	0.97
4	5.82	5.5	5.59	79.67	93.5	34676.78	0.94
Total	4.78	4.28	4.27	71.26	79.86	16195.27	**0.94**

Another result of the analyzes is the distribution of the efficient countries to the clusters. Some countries became efficient with low HDI and WEF scores (the countries in the second cluster), whereas some others are efficient with high HDI and WEF scores (the countries in the third and fourth clusters). This discrimination may show more specific path for the development of other inefficient or low-scored countries. Such that countries in third or fourth cluster should follow the efficient countries in their own clusters to get better, while the countries in first or second clusters should develop in the way of the efficient countries in their clusters. More can be offered to the efficient countries in the second cluster. These countries are efficient according to competitiveness and human development perspectives, but they do not have high scores in these indicator. Thus, they should follow their balanced development of both perspectives in order to reach the level of third cluster countries.

6. CONCLUSIONS AND FURTHER SUGGESTIONS

In this study, the capability of countries to convert their competitiveness power into better lives for their people has been investigated through a two-stage study based on DEA and ANN analysis. In the first stage, WEF scores related to BASREQ, EFFENH, and INSOPF were used as the inputs of the DEA, while the HDI scores were taken as the outputs. The results of the DEA gave super-efficiency scores of the countries for three years. In addition to this analysis, a cluster analysis is conducted utilizing the WEF and HDI scores that were used in the DEA. SOMs are used to determine the number of clusters, where starting from eight clusters, it is found that four clusters are good for these data. The countries are grouped under four clusters where higher cluster number indicates more successful countries in terms of WEF and HDI scores.

By utilizing the DEA in this study, the efficiency of converting competitiveness of countries to the development of their citizens are analyzed. As a result of the DEA it is found that some countries with high competitiveness levels do not have such high human development levels relatively (e.g., Germany and Japan). On the contrary, some countries with low scores of competitiveness have found to be efficient which means that they have high relatively high human development levels (e.g., Argentina and Venezuela).

However, the efficiency scores, solely, may not be a sound reference when the differences of development of the countries are not taken into account. For this purpose, cluster analysis is conducted in this study to expose the development level of the countries and state the relationship between the development and efficiency. According to the ANN cluster analysis results, there are four development levels for the countries considered. It is concluded that the efficiencies of the countries in different clusters have a different meaning at all.

The results of both DEA and cluster analysis are generally robust and stable for the evaluation of different years. Such that efficiency scores of seven countries of 44 are changed during the three years evaluation. Additionally only clusters of four countries are changed in the same period. These results are logical when it is accepted that HDI or WEF scores would not change dramatically in three years.

Further improvement of this work could be a sector-based evaluation among the countries or the comparison of different sectors within a country. The use of the three dimensions of the WEF competitiveness index as the three inputs assumes that all sectors within a nation are at the same stage concurrently. However, this is not the case in the countries where some sectors are more competitive than others in global market.

The research conducted could also be improved further by putting lower and upper bounds on some of the inputs, as it may not be very realistic to believe that an infinite reduction or an unlimited increase could be realized in some of the attributes in an attempt to generate a greater efficiency score for a country. Finally, imposing ratios between certain inputs may also be more realistic.

In this study, the dynamic structure of the problem was not taken into account. However, the connection between human development and competitiveness is a two-way interaction. For example, education is an important contributor to technological capability and technical change in industry. Similarly, improved health has direct effects on labor productivity, especially among poorer countries. Education has an impact on the nature and growth of exports, which, in turn, affects the aggregate growth rate. This is another way in which human development influences macroeconomic performance. Therefore, a dynamic model with feedback to analyze these interrelationships through time lags would be useful.

ACKNOWLEDGMENT

The authors are grateful to REF-The TUSIAD Sabanci University Competitiveness Forum for their help in gathering the data.

REFERENCES

Anand, S., & Sen, A. (2000). Human development and economic sustainability. *World Development*, *28*(12), 2029–2049.

Andersen, P., & Petersen, N. C. (1993). A procedure for ranking efficient units in data envelopment analysis. *Management Science*, *39*, 1261–1264.

Artto, E. W. (1987). Relative total costs-an approach to competitiveness measurement of industries. *Management International Review*, *27*, 47–58.

Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale efficiencies in data envelopment analysis. *Management Science*, *30*, 1078–1092.

Blanke, J., & Lopez-Claros, A. (2004). The growth competitiveness index: Assessing countries' potential for sustained economic growth. In: WEF, *The Global Competitiveness Report 2004–2005*. New York: Oxford University Press for the World Economic Forum.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, *2*(6), 429–444.

Cherchye, L., Moesen, W., Rogge, N., Van Puyenbroeck, T., Saisana, M., Saltelli, A., Liska, R., & Tarantola, S. (2008). Creating composite indicators with DEA and robustness analysis: the case of the technology Achievement index. *Journal of Operational Research Society*, *59*, 239–251.

Cooper, W. W., Seiford, L. M., & Tone, K. (2000). *Data envelopment analysis: A comprehensive text with models, applications, references and DEA-solver software*. Boston: Kluwer Academic Publishers.

Costantini, V., & Monni, S. (2007). Environment, human development and economic growth. *Ecological Economics*, *64*(4), 867–880.

Cubbin, J., & Tzanidakis, G. (1998). Regression versus data envelopment analysis for efficiency measurement: An application to the England and Wales regulated water industry. *Utilities Policy*, *7*(2), 75–85.

Davies, A., & Quinlivan, G. (2006). A panel data analysis on the impact of trade on human development. *Journal of Socio-Economics*, *35*(5), 868–876.

Despotis, D. K. (2005a). Measuring human development via data envelopment analysis: The case of Asia and the Pacific. *Omega*, *33*, 385–390.

Despotis, D. K. (2005b). A reassessment of the HDI via data envelopment analysis. *Journal of Operational Research Society*, *56*, 969–980.

Dunning, J. H. (1990). The globalization of firms and the competitiveness of countries. In: J. H. Dunning, B. Kogut & M. Blomström (Eds), *Globalization of firms and the competitiveness of nations* (pp. 9–57). Lund: Lund University Press.

Dunning, J. H. (1993). *The globalization of business: The challenge of the 1990s*. London: Routledge.

Dyson, R. G., Allena, R., Camanhob, A. S., Podinovskia, V. V., Sarricoa, C. S., & Shalea, E. A. (2001). Pitfalls and protocols in DEA. *European Journal of Operational Research*, *132*(2), 245–259.

Fried, H. O., Lovell, C. A. K., & Schmidt, S. S. (1993). *The measurement of productive efficiency-techniques and applications*. New York, NY: Oxford University Press.

Gan, G., Ma, C., Wu, J. (2007). Data clustering: Theory, algorithms, and applications. In: *ASA-SIAM series on statistics and applied probability* (illustrated edition). Philadelphia, PA: SIAM, Society for Industrial and Applied Mathematics.

Gattoufi, S., Oral, M., Kumar, A., & Reisman, A. (2004). Content analysis of data envelopment analysis literature and its comparison with that of other OR/MS fields. *Journal of Operational Research Society*, *55*, 911–935.

Golany, B., & Thore, S. (1997). The economic and social performance of nations: Efficiency and returns to scale. *Socio-Economic Planning Sciences*, *31*(3), 191–204.

Hair, J., Anderson, K. E., & Black, W. C. (1995). *Multivariate data analysis with readings*. New York: Prentice Hall.

Hawksworth, J. (2006). *The World in 2050: How big will the major emerging market economies get and how can the OECD compete?* London, England: PricewaterhouseCoopers. Available at http://www.pwc.com/en_GX/gx/world-2050/pdf/world2050emergingeconomies.pdf

http://hdr.undp.org/hd/

http://www.imd.ch

http://www.un.org.my/uploads/files/HDI_Technical_note.pdf

Kogut, B. (1988). Country patterns in international competition: Appropriability and oligopolistic agreement. In: N. Hood & J. E. Vahlne (Eds), *Strategies in global competition*. New York: Wiley.

Kohonen, T. (1987). Adaptive associative and self-organizing functions in neural computing. *Applied Optics*, *26*(23), 4910–4918.

Lall, S. (2001). Competitiveness indices and developing countries: An economic evaluation of the global competitiveness report. *World Development*, *29*, 1501–1525.

Lau, K-N., & Lam, P-Y. (2002). Economic freedom ranking of 161 countries in year 2000: A minimum disagreement approach. *Journal of Operational Research Society*, *53*, 664–671.

Li, S., Jahanshahloo, G. R., & Khodabakhshi, M. (2007). A super-efficiency model for ranking efficient units in data envelopment analysis. *Applied Mathematics and Computation*, *184*(2), 638–648.

Lovell, C. A. K., & Rouse, A. P. B. (2003). Equivalent standard DEA models to provide super-efficiency scores. *Journal of Operational Research Society*, *54*, 101–108.

Mangiameli, P., Chen, S. K., & West, D. A. (1996). Comparison of SOM neural network and hierarchical clustering. *European Journal of Operational Research*, *93*(2), 402–417.

McArthur, J. W., & Sachs, J. D. (2001). The growth competitiveness index: measuring technological advancement and the stages of development. In: WEF, *The Global Competitiveness Report 2001–2002*. New York: Oxford University Press for the World Economic Forum.

Milligan, G. W. (1980). An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrica*, *45*(3), 325–342.

Onsel, S., Ulengin, F., & Ulengin, B. (2004). A dynamic approach to scenario analysis: The case of Turkey's inflation estimation. *European Journal of Operational Research*, *158*(1), 124–145.

Porter, M. E. (1990). *The competitive advantage of nations*. London: Macmillan.

Ranis, G., Stewart, F., & Ramirez, A. (2000). Economic growth and human development. *World Development*, *28*(2), 197–219.

Retzlaff-Roberts, D., Chang, C. F., & Rubin, R. M. (2004). Technical efficiency in the use of health care resources: A comparison of OECD countries. *Health Policy*, *69*(1), 55–72.

Sagar, A. D., & Najam, A. (1998). The human development index: A critical review. *Ecological Economics*, *25*, 249–264.

Sala-i-Martin, X., & Artadi, E. V. (2004). The global competitiveness index. In: WEF, *The Global Competitiveness Report 2004–2005*. New York: Oxford University Press for the World Economic Forum.

Seiford, L. M., & Zhu, J. (1998). An acceptance system decision rule with data envelopment analysis. *Computers and Operations Research*, *25*(4), 329–332.

UNDP (United Nations Development Program). (2006). Human Development Report 2005: International cooperation at a crossroads: Aid, trade and security in an unequal world. Available at http://hdr.undp.org/en/reports/global/hdr2005/

UNDP (United Nations Development Program) (2007). Human Development Report 2006: Beyond scarcity: power, poverty, and global water crisis. Available at http://hdr.undp.org/en/reports/global/hdr2006/

UNDP (United Nations Development Program) (2008). Human Development Report 2007/2008: Fighting climate change: Human solidarity in a divided world. Available at http://hdr.undp.org/en/reports/global/hdr2007–2008/

UNDP (United Nations Development Program). (1990). *Human development report 1990*. New York: Oxford University Press.

WEF (World Economic Forum). (2004). *The global competitiveness report, 2004–2005, September 2005*. Geneva, SW: Palgrave Macmillan.

WEF (World Economic Forum). (2005). *The global competitiveness report, 2005–2006, September 2005*. Geneva, SW: Palgrave Macmillan.

WEF (World Economic Forum). (2006). *The global competitiveness report, 2006–2007, September 2006*. Geneva, SW: Palgrave Macmillan.

WEF (World Economic Forum). (2007). *The global competitiveness report, 2007–2008, September 2006*. Geneva, SW: Palgrave Macmillan.

Zanakis, S. H., & Becerra-Fernandez, I. (2005). Competitiveness of nations: A knowledge discovery examination. *European Journal of Operational Research*, *166*(1), 185–211.

RICH AND POOR IN SAINT LOUIS: PERFORMANCE CHARACTERISTICS OF PUBLIC SCHOOLS USING A DATA ENVELOPMENT ANALYSIS APPROACH

N. K. Kwak and Walter A. Garrett, Jr.

ABSTRACT

Many urban areas of the United States have experienced urban sprawl in the past 60 years. Severe out-migration of relatively wealthier families to ex-urban counties has left relatively poorer families behind. When combined with the recent national conversation about school improvement, this migration has caused significant stress on urban school districts, as indicated by population demographics, revenues, and school performance.

This chapter looks at 22 public school districts in Saint Louis County, Missouri. It first reviews the decision environment for those districts and constructs a relative wealth variable from environmental factors. Then, using data envelopment analysis (DEA), it compares rich districts and poor districts, and attempts to classify the relative efficiencies of those districts. Three DEA models are considered: the baseline CCR-O

Financial Modeling Applications and Data Envelopment Applications
Applications of Management Science, Volume 13, 227–246

ISSN: 0276-8976/doi:10.1108/S0276-8976(2009)0000013014

(Model 1), a CAT-O-C model (Model 2), and a revised CCR-O (Model 3). Using computer software, DEA-Solver, these three model results are compared and analyzed to study the effects of each district's relative wealth on the model results.

The study concludes that adding a relative wealth variable produces more robust model results and suggests that school district decisions may be improved by including a relative wealth variable in their decision-making processes.

1. INTRODUCTION

Saint Louis County, Missouri, is at the crossroads of a continent, a few miles from the population center of the United States and, in many ways, typical of suburban life in America.

It has 24 public school districts, each providing service to a geographic portion of the County. Some provide excellent outcomes; some struggle to remain accredited. Some are rich, with enough resources to fund the very highest levels of education quality with money left over; some are so poor they ration chalk in the classrooms.

Rich or poor, excellent or not, all these districts are judged by a uniform set of standards. The state's primary "funding formula" that is used to apportion state money to local school districts disregards variation of all resources except local tax revenues. Similarly, the state's assessment and accreditation process applies a uniform set of standards to all local districts, no matter what their circumstances.

Our study begins with the premise that outcomes depend, at least in part, on environmental factors. Section 2 describes the historical and political environment in which Saint Louis County school districts operate.

Section 3 describes the construction of a rich-and-poor variable, which we use as a proxy for environmental factors such factors being notoriously difficult to measure and model. The rich-and-poor variable is then used to categorize local school districts for further analysis.

We employ the proven technology of data envelopment analysis (DEA) and construct a baseline model for 22 school districts in Saint Louis County. Then we modify that model using the rich-and-poor variable and run two additional models using that additional variable. Section 3 describes the DEA models and the variables used in the analysis.

Section 4 briefly describes the results of the models and introduces conclusions to be drawn from the models. Section 5 offers some concluding remarks about this study and offers suggestions for further study of this topic.

2. PERFORMANCE ENVIRONMENT

Managers in Missouri public school districts face a complex resource allocation problem with numerous constraints. On a simplistic level, they receive tax revenues and spend them to educate children. But planning becomes complicated when trying to predict what level of revenues will be received, how many children will take seats, and what level of instruction will be needed for every individual child. Schools exist in a dynamic environment where every year is likely to be different from the previous one. The resource allocation issues these managers face may be broadly described as those involving population migration and demographic shifts, education reform and improvement, and funding equity and adequacy.

2.1. Demography

Saint Louis, Missouri, a city known as "The Gateway to the West," has always known population migration. The city and its environs were the pre-Columbian home to tens of thousands of Mississippian mound-builders, who either vanished or were displaced by European explorers and traders. The modern City of Saint Louis was born as a French trading outpost in the 18th century. It flourished as a primary commercial, transportation, and military center during America's westward expansion of the 19th century. Massive immigration of English, Irish, and German settlers helped Saint Louis grow to become America's fourth-largest city by 1900. In the 20th century, population growth was aided by the great northern migration of African Americans from southern states. By the 1950 census, the City of Saint Louis had reached its peak population of about 857,000.

After 1950, St. Louis immigrants became emigrants. Saint Louis City had separated itself from Saint Louis County in 1876. Then largely rural, the county area (to the north, west, and south of Saint Louis) slowly grew in population. After World War II, emigration to Saint Louis County accelerated. The County, not the City, was the primary beneficiary of the post-war baby boom. Cities, businesses, and community institutions of every kind developed in Saint Louis County, whereas the population of the City of Saint Louis suffered a precipitous decline. In the 2000 census, City population numbered below 350,000. County population was then slightly more than 1,000,000.

The population growth and decline of the City of Saint Louis have been repeated in Saint Louis County. The County is now (2009) experiencing its own emigration and population appears to have peaked. New cities in counties to the west and south are now experiencing their own growth booms.

2.2. School Districts

As Saint Louis County developed, it was organized into 23 public school districts, based on political and geographic boundaries.[1] These districts are independent of any county or city government; each district is responsible for raising its own local revenues and managing its own affairs. Districts are supervised at the state level by Missouri's Department of Elementary and Secondary Education (DESE).

Each of these school districts has experienced its own population shift and the consequent challenge to planning and management. But migration of population is only one of several demographic issues facing the districts. Partly because of their development history, the school district in the City of Saint Louis and most of the County districts were involved in court-ordered desegregation programs for almost three decades, beginning in the 1970s. Desegregation brought new students and new funding to county districts, along with new cultural and socio-economic problems; elimination of the desegregation programs reduced student populations and funding.

2.3. Education Reform

Concurrently with the population changes the districts were experiencing, there began an accelerating national movement for school reform. The "A Nation at Risk" report (National Commission on Excellence in Education, 1983) spawned national and local programs to improve school performance. In Missouri, DESE implemented the Missouri School Improvement Program (MSIP), which serves as a master plan for setting and reviewing educational performance standards and, ultimately, serves as the basis for accreditation of local school districts. A key component of MSIP is the Missouri Assessment Program (MAP), a series of standardized tests administered to students at key grade levels. If successful, MSIP and MAP serve as both the impetus for school reform and a source of data by which improvement can be measured.

At the national level, responsibility for measuring school improvement (and collecting the necessary data) was vested in the National Assessment of Educational Progress (NAEP), within the National Center for Education Statistics (NCES), a division of the U.S. Department of Education. NAEP collects and publishes data using a standard format across all public school districts in the United States.

The national education reform movement produced the No Child Left Behind Act (NCLB) in 2001. While allowing states (and local districts) to develop and administer their own standards of performance, NCLB mandates continuous performance improvement. Each local school and district is required to make Adequate Yearly Progress (AYP) toward its improvement goals and imposes sanctions against those which do not.

2.4. Funding

Finally, districts face problems of funding equity and adequacy. The Missouri state Constitution requires the state legislature to "establish and maintain free public schools for the gratuitous instruction of all persons in this state…." But a series of law suits since 1993 have argued that the state does not adequately nor equitably fund local districts. In the most recent case, almost half the districts in the state joined a suit that is now awaiting a decision by the Missouri Supreme Court. More than 100 (of the state's 523 districts) receive no money from the state's basic funding formula, relying instead on federal funds and local taxes. (An excellent review of the issues and history related to Missouri school funding is in Welker, 2006.)

Even if state monies were equitably distributed to local districts, a question of adequacy remains. Under the present (2009) economic crisis, the state government is experiencing revenue shortfalls. That means the amount of money the state actually has available for distribution to local districts might be far less than the state intends to give. As of January 2009, Missouri was projecting a $342 million revenue shortfall (Dunson, 2009).

2.5. Environmental Significance

Such is the environment within which school district managers make decisions. Environmental issues are exogenous to the decision process; that is, these issues are totally beyond the control of the decision maker. But in a multi-criteria decision process, whatever the decision technology, these issues act as constraints – very real constraints, although difficult to quantify.

A defining characteristic of this environment is that it is rapidly dynamic: fluctuating populations, migrating social groups, revised performance goals, uncertain annual funding. But the school district's boundaries are static against this fluid background; the district cannot move when its families move and it cannot market itself to attract a different customer. Thus, yesterday's rich district may be left behind by its changing environment to become tomorrow's poor district.

These environmental interpretations suggest the need for a decision process that explicitly or implicitly accommodates the rich-and-poor realities of local districts. The process either must be dynamic or have a short-term orientation that permits rapid adaptation to its changing environment. A funding formula or a performance goal that worked very well in past year's rich district environment may be totally inadequate in the changed reality of this year's poor district.

To a great extent, no school district chooses to be rich or poor; it becomes so by a naturally occurring environmental process. The complex interactions of environmental forces, therefore, culminate in a district being rich or poor. If that is so, then a single measure or indicator of relative district wealth (a "rich-and-poor" variable) could serve as a valuable proxy for these interrelated environmental factors. We explore this possibility in Section 3.

3. MODEL BACKGROUND AND DEVELOPMENT

For this chapter, we analyzed 22 public school districts in Saint Louis County, Missouri. Those districts represent a good (although non-random) sample of the 523 districts in Missouri. Geographically, some are very small and others are very large. Some have very small enrollments and others have very large. Some are among the poorest districts in Missouri; others are the richest. On any scale of performance, Saint Louis County has districts at both extremes. Saint Louis County, therefore, serves as an introduction to further analyses and interpretations that might affect the entire state.

3.1. Previous Studies

The issues addressed in this chapter are not unique to Missouri and Saint Louis County. The literature of the school improvement movement contains a plethora of measurement and evaluation methods. Prominent among these is the method of DEA. In their original paper, Charnes, Cooper, and

Rhodes (1978) describe the model as a tool for measuring the relative efficiencies of production units, called decision-making units (DMUs). DEA uses the underlying method of linear programming (LP) to segregate "efficient" DMUs from "inefficient" ones and also identifies how much improvement each inefficient DMU needs to become efficient. DEA thus focuses on both the inputs and the outputs of a process and identifies the relative changes in either, which may lead to improved efficiencies. [For a detailed description of the DEA methodology, see Cooper, Seiford, and Tone (2007), Kwak and Kim (2000), or Kwak and Lee (2009).]

Numerous extensions and applications of the DEA model have followed its 1978 introduction. The "production" and "efficiency" focus of DEA makes it directly applicable to the decision environment of public school districts, and particularly attractive to the language of school reform. The method was quickly applied in DEA studies of Texas elementary schools (Bessent and Bessent, 1980) and community colleges (Bessent and Bessent, 1981). New York schools were studied using DEA by Ruggiero and Vitaliano (1999). Failing Missouri schools were studied using DEA by Primont and Domazlicky (2006). Numerous other education applications are found in the DEA literature.

To address the question of equity in Missouri's 2005 revision to its state funding formula, Welker (2006) studied the relationship between the revised school funding formula and district performance. Instead of using DEA, he used correlations and hypothesis testing, and concluded that the funding formula was inequitable. That study addressed the central issue now being decided by the Missouri Supreme Court (Lieb, 2009).

3.2. Three DEA Models

This chapter uses DEA to study 22 of the 24 independent public school districts in Saint Louis County. (One district was omitted for lack of data; another was omitted because of its unique characteristics.) Using an a priori designation to categorize the relative wealth of each district, it attempts to determine whether the solution of the basic DEA model can be improved by the addition of the wealth variable. If so, district wealth should be considered when attempting efficiency improvements. We present three variations of the DEA model and compare the results to each other.

The "basic" DEA model is the original 1978 model (Charnes et al., 1978), called "CCR" after its developers. It exists in two variants, the CCR-I and the CCR-O, which focus on input improvements and output improvements, respectively. For purposes of this chapter, the CCR-O model is considered the "baseline" ("Model 1"). DEA-Solver, a flexible computer program developed by Tone and bundled with the Cooper et al. (2007) text book, was used to solve all DEA models used for this chapter.

An extension to the basic model, called CAT, allows for special treatment of categorical variables. In the CAT model, DMUs are divided into hierarchical subsets or categories determined by some qualitative characteristic which is shared by DMUs within each category, but is different from the other categories. An additional variable is used to explicitly relate each DMU to its membership in one mutually exclusive category. Starting with the lowest-numbered category (the most severely "handicapped"), DMUs are compared only to other DMUs within that category, not to the entire data set. As the model moves progressively up the categorical hierarchy, DMUs in each category are compared to other DMUs both within that category and in all lower categories (but never to DMUs in higher categories).

The significance of the CAT approach is that a DMU which might be "inefficient" as compared to all DMUs in the data set may in fact be "efficient" when taking its handicap into consideration. [See Cooper et al. (2007) for a complete treatment of this method.] In this chapter, the categorical model is employed to identify relative district wealth ("rich and poor") and take that measurement into account using the DEA method ("Model 2").

Conceptually, the categorical variable functions somewhat like a "dummy" variable, employed frequently in ordinary least squares (OLS) models. The dummy or constructed variable marks the presence or absence of a binary data attribute or represents a scalar value of a discrete attribute. As an extended test of the model, we therefore went back to the basic model and used the categorical variable as an additional input variable in the CCR-O formulation ("Model 3").

Data used in this chapter were derived from the DESE online data base. Various reports for each district in Saint Louis County were downloaded and merged; the data set was then reduced to the applicable variables. All data represent the 2007–2008 school year, which were reported to DESE by individual districts in the fall of 2008, and were downloaded from the DESE web site in the spring of 2009.

The input variables selected for this study were one financial measure (CUREXP, current district expenditures in dollars) and two measures of staff resources (TEACHSAL, average teacher salary in dollars, and NUMTEACH, number of classroom teachers adjusted for full-time equivalence in the classroom). One output variable addresses the quantity of education delivered (ADA, average daily attendance). Three additional

Table 1. Model Variables.

Type	Variable	Description	Model
Input (district resources)	CUREXP	Total current expenditures in dollars	1,2,3
	TEACHSAL	Average teacher salaries in dollars	1,2,3
	NUMTEACH	Number of classroom teachers (full-time equivalents)	1,2,3
	3 = RICH	Relative district wealth (see "Categorical", below)	3
Output (district outcomes)	ADA	Average daily attendance (number of students)	1,2,3
	AYP	Adequate yearly progress (against annual goals)	1,2,3
	ACT	Average composite score of students taking the American College Test (ACT)	1,2,3
	GRAD	Graduation rate (percentage) of the student cohort entering four years ago	1,2,3
Categorical (CAT model)	3 = RICH	Relative district wealth	2

Note: See Table 3 for disaggregation based on:
(a) Current annual expenditures total assessed valuation.
(b) Percentage of students receiving free or reduced-price lunch.
(c) Percentage of minority race students.

output variables address the quality of education delivered: AYP, a measure of adequate yearly progress; ACT, the average composite ACT score of students who took the ACT exam; and GRAD, the graduation rate in percent of the student cohort which entered the school four years earlier.

Table 1 summarizes the definitions of all variables. Table 2 lists the detailed values of each variable.

3.3. Rich-and-Poor Variable

The rich-and-poor categorical variable (3 = RICH) used in Models 2 and 3 was constructed using a decision rule applied to three environmental variables. The first measure is current expenditures as a percentage of total assessed valuation. Because most local revenues are derived from property taxes, this variable measures the district's need (and will) to tax property at a high rate; higher rates are construed to indicate financially needy ("poor") districts. The second measure is the percentage of students receiving free or

Table 2. DMUs and Variable Values.

District	DMU	Input Variables				Output Variables			
		CUREXP, $	TEACHSAL, $	NUMTEACH	3 = RICH	ADA	AYP	ACT	GRAD
Affton	AFT	21,538,856	54,259	132	2	2,270.2	13	22.5	91.0
Bayless	BAY	10,407,172	42,963	83	2	1,558.2	12	20.6	90.0
Brentwood	BRN	11,205,067	62,599	66	2	723.3	14	22.7	95.7
Clayton	CLA	39,042,872	65,682	207	3	2,345.4	14	25.4	97.4
Ferguson-Florissant	FFL	115,180,759	55,044	717	1	11,191.0	9	18	93.0
Hancock Place	HAN	15,637,410	47,356	113	1	1,728.1	10	20	83.6
Hazelwood	HAZ	172,970,748	50,045	1,127	1	18,180.6	10	19.1	84.5
Jennings	JEN	34,986,591	59,685	185	1	3,112.3	5	17.4	84.1
Kirkwood	KRK	50,343,445	59,138	293	3	4,491.5	13	23.5	93.5
Ladue	LAD	45,043,371	58,672	275	3	3,448.1	14	25.9	97.3
Lindbergh	LIN	48,929,436	53,082	311	3	5,211.6	14	23.5	95.2
Mehlville	MEL	77,693,539	51,505	578	2	10,139.5	14	21.5	91.1
Maplewood-Richmond Hts	MRH	12,882,798	49,007	75	2	952.6	11	19.1	88.6
Normandy	NOR	53,289,581	57,977	289	1	3,915.8	4	16.4	61.1
Parkway	PAR	177,889,950	55,176	1,061	3	16,685.0	13	24.1	93.2
Pattonville	PAT	66,550,028	57,263	362	2	5,521.2	13	22.1	86.8
Ritenour	RIT	50,606,196	55,524	335	2	5,880.0	9	19.9	87.8
Riverview Gardens	RIV	59,566,851	53,037	368	1	6,411.7	1	17	65.6
Rockwood	ROK	175,059,783	50,951	1,245	3	20,641.1	14	24.2	95.8
University City	UCT	37,162,478	53,371	205	2	3,052.5	7	18.7	80.9
Valley Park	VPK	9,681,719	51,210	62	2	950.3	14	20.9	94.9
Webster Groves	WEB	40,178,466	57,885	258	2	3,789.2	14	23.4	97.9
	Mean	60,265,778	54,611	379	2	6,009.0	11.0	21.2	88.6
Data Summary	Minimum	9,681,719	42,963	62	1	723.3	1.0	16.4	61.1
	Maximum	177,889,950	65,682	1,245	3	20,641.1	14.0	25.9	97.9

reduced-price lunch at school. This measure has long been used in the education reform literature as a proxy for family poverty; the higher the rate, the poorer the district. The third measure is the percentage of student population in minority racial groups; the higher the rate, the poorer the district.

Taken individually, any (or all) of these three wealth measures could trigger endless debates on the proper interpretation of the measure. We avoid that problem using a simple decision process: First, rank each of the three variables and divide the 22 DMUs into quartiles for each variable. Second, inspect the quartiles thus obtained. If a DMU appears in the top quartile for any two out of the three ranked variables, call it "rich." If a district appears in the bottom quartile for any two of the three ranked variables, call it "poor." Call the remaining districts "average." These labels are then converted to a numeric variable (3 = RICH) where 1 = poor (the most severely handicapped, in the language of Cooper et al.), 2 = average, and 3 = rich. Of the 22 DMUs in the study, six were deemed to be poor and six were deemed to be rich; ten were average.

Table 3 shows the variables and values used to construct the rich-and-poor categorical variable.

4. MODEL ANALYSIS AND DISCUSSION

So that we could assess the benefit of including the rich-and-poor categorical variable, we compared three models: the baseline CCR-O (Model 1), a CAT-O-C model (Model 2), and a revised CCR-O model which includes the categorical variable as an input (Model 3).

4.1. Model 1

Results of the baseline CCR-O model are summarized in Table 4. In column 2, a score of 1 indicates that the DMU (school district) is operating efficiently. A score of less than 1 indicates that the district is inefficient, relative to all the other districts being evaluated. Five of the 22 districts were found to be efficient. Thus, these five districts are interpreted to produce a higher combination of outputs for any given level of inputs (resources).

In column 4, each inefficient district is referenced to one or more efficient districts that have similar input/output characteristics to the inefficient district. These referenced (efficient) districts help to highlight an inefficient

Table 3. Construction of Rich-and-Poor Variable (3 = RICH).

DMU	Current Expenditures (CE), $	Assessed Value (AV), $	CE/AV Pct. (a)	Lunch Pct. (b)	Min'ty Pct. (c)	(a) Rank	(b) Rank	(c) Rank	Times in 1 Qt	Times in 4 Qt	3 = RICH
AFT	21,538,856	432,748,290	5.0	29.5	14.9	12	13	20			2
BAY	10,407,172	175,541,500	5.9	51.8	22.0	9	9	18		1	2
BRN	11,205,067	300,333,050	3.7	21.5	30.6	19	14	12		1	2
CLA	39,042,872	1,053,238,080	3.7	13.9	33.7	20	20	11		2	3
FFL	115,180,759	1,122,602,360	10.3	62.7	79.4	5	6	5	3		1
HAN	15,637,410	73,984,430	21.1	70.5	29.6	3	4	14	2		1
HAZ	172,970,748	2,053,908,400	8.4	46.1	69.3	6	11	6	2		1
JEN	34,986,591	119,509,900	29.3	81.2	98.8	1	2	2	3		1
KRK	50,343,445	1,251,226,290	4.0	16.2	24.9	17	18	17		3	3
LAD	45,043,371	1,526,027,185	3.0	7.7	28.5	22	22	15		2	3
LIN	48,929,436	1,334,552,970	3.7	14.0	12.5	21	19	22		3	3
MEL	77,693,539	1,864,499,620	4.2	19.8	14.9	16	15	21			2
MRH	12,882,798	286,862,900	4.5	52.6	45.7	14	8	8			2
NOR	53,289,581	275,162,170	19.4	82.6	99.2	4	1	1	3		1
PAR	177,889,950	4,631,201,990	3.8	17.0	30.0	18	17	13		2	3
PAT	66,550,028	1,490,383,180	4.5	35.1	36.5	15	12	10			2
RIT	50,606,196	638,489,100	7.9	62.8	51.5	7	5	7	1		2
RIV	59,566,851	275,772,230	21.6	78.9	98.1	2	3	3	3		1
ROK	175,059,783	3,589,547,265	4.9	13.0	17.7	13	21	19		2	3
UCT	37,162,478	638,919,200	5.8	59.1	88.4	10	7	4			2
VPK	9,681,719	157,613,050	6.1	46.7	37.4	8	10	9			2
WEB	40,178,466	756,220,120	5.3	19.5	27.3	11	16	16			2

Table 4. Model 1 (Baseline, CCR-O) Solutions of Efficiency Score and Reference Set.

DMU	Score	Rank	Reference Set (DMUs and Intensities)						#Ref[a]
AFT	0.930	11	BAY	1.194	MEL	0.057			
BAY	1.000	1	BAY	1.000					16
BRN	1.000	1	BRN	1.000					0
CLA	0.807	21	BAY	1.496	ROK	0.028			
FFL	0.912	15	BAY	0.212	MEL	0.622	ROK	0.273	
HAN	0.881	16	BAY	1.087	ROK	0.013			
HAZ	0.969	8	MEL	0.141	ROK	0.840			
JEN	0.927	12	BAY	1.216	MEL	0.144			
KRK	0.862	18	BAY	0.952	MEL	0.341	ROK	0.013	
LAD	0.921	14	BAY	1.264	ROK	0.086			
LIN	0.956	9	BAY	0.833	MEL	0.266	ROK	0.070	
MEL	1.000	1	MEL	1.000					12
MRH	0.923	13	BAY	0.494	VPK	0.543			
NOR	0.759	22	BAY	0.907	MEL	0.369			
PAR	0.941	10	BAY	0.148	MEL	0.213	ROK	0.743	
PAT	0.860	19	BAY	0.704	MEL	0.525			
RIT	0.990	6	BAY	0.723	MEL	0.475			
RIV	0.984	7	BAY	0.569	MEL	0.555			
ROK	1.000	1	ROK	1.000					9
UCT	0.825	20	BAY	0.987	MEL	0.213			
VPK	1.000	1	VPK	1.000					1
WEB	0.867	17	BAY	1.204	ROK	0.121			

[a]#Ref: number of times this efficient DMU is a reference for an inefficient DMU.

district's weaknesses. The comparison shows the relative performance of the inefficient unit compared with its closest peers (its reference set). For example, the inefficient Pattonville District (PAT) is referenced to the efficient Mehlville (MEL) and Bayless (BAY) districts. Note that reference sets may be different for each inefficient district.

Column 10 of Table 4 contains a single number for each efficient district. That number represents the number of times the efficient district is referenced to an inefficient district. For example, the efficient Rockwood district (ROK) appears nine times (in column 4) as a reference district for nine different inefficient districts.

Table 5 presents the sensitivities of each DMU to its input and output variables. Input slack ("excess") indicates that inefficient districts could reduce their inputs by the indicated amounts without adversely affecting outputs. Output slack ("shortage") indicates that inefficient districts should be able to increase their outputs by the indicated amounts without requiring additional inputs. For example, the Clayton District (CLA) has one of the

Table 5. Model 1 (Baseline, CCR-O) Sensitivity Analysis for Input/Output Variables.

DMU	Excess CUREXP, $	Excess TEACHSAL, $	Excess NUMTEACH	Shortage ADA	Shortage AYP	Shortage ACT	Shortage GRAD
AFT	4,669,425	0	0.0	0.0	1.2	1.6	14.9
BAY	0	0	0.0	0.0	0.0	0.0	0.0
BRN	0	0	0.0	0.0	0.0	0.0	0.0
CLA	18,583,798	0	47.3	0.0	1.0	0.0	16.6
FFL	16,917,225	0	0.0	0.0	5.2	4.6	0.0
HAN	2,051,627	0	6.0	0.0	1.9	0.0	4.1
HAZ	15,017,589	0	0.0	0.0	3.4	3.6	6.1
JEN	11,114,121	0	0.0	0.0	11.2	9.4	31.8
KRK	11,656,190	0	0.0	0.0	1.3	0.0	9.5
LAD	16,849,487	0	63.2	0.0	1.2	0.0	16.4
LIN	7,237,307	0	0.0	0.0	0.1	0.0	6.4
MEL	0	0	0.0	0.0	0.0	0.0	0.0
MRH	2,489,246	0	0.0	253.6	1.6	0.8	0.0
NOR	15,148,992	0	0.0	0.0	10.8	5.0	34.7
PAR	29,724,871	0	0.0	0.0	1.3	0.0	4.9
PAT	18,460,503	0	0.0	0.0	0.7	0.1	10.3
RIT	6,209,094	0	0.0	0.0	6.2	5.0	19.7
RIV	10,519,290	0	0.0	0.0	13.6	6.4	35.1
ROK	0	0	0.0	0.0	0.0	0.0	0.0
UCT	10,336,717	0	0.0	0.0	6.3	2.3	10.2
VPK	0	0	0.0	0.0	0.0	0.0	0.0
WEB	6,510,120	0	7.4	0.0	0.0	0.8	7.1

smallest class sizes in the County; this table suggests CLA should be able to reduce the number of classroom teachers by 47 without adversely affecting performance.

4.2. Model 2

Results of the CAT model are summarized in Table 6. This model found nine efficient districts, four more than Model 1. Notably, all four are poor. When compared to districts within their own wealth category, they are efficient: Hancock Place (HAN), Hazelwood (HAZ), Jennings (JEN), and Riverview Gardens (RIV). Note that only one of the nine, Rockwood (ROK), is rich.

The remaining districts are still identified as inefficient. Their scores are essentially unchanged, but their reference sets have changed. The four poor districts newly identified as efficient can now be used as references for

Table 6. Model 2 (CAT-O-C) Solutions of Efficiency Score and Reference Set.

DMU	Score	Rank	Reference Set (DMUs and Intensities)						#Ref[a]
AFT	0.930	14	BAY	1.194	MEL	0.057			
BAY	1.000	1	BAY	1.000					11
BRN	1.000	1	BRN	1.000					0
CLA	0.807	21	BAY	1.496	ROK	0.028			
FFL	0.979	11	HAN	0.553	HAZ	0.576			
HAN	1.000	1	HAN	1.000					2
HAZ	1.000	1	HAZ	1.000					1
JEN	1.000	1	JEN	1.000					1
KRK	0.862	18	BAY	0.952	MEL	0.341	ROK	0.013	
LAD	0.921	16	BAY	1.264	ROK	0.086			
LIN	0.956	12	BAY	0.833	MEL	0.266	ROK	0.070	
MEL	1.000	1	MEL	1.000					8
MRH	0.923	15	BAY	0.494	VPK	0.543			
NOR	0.796	22	HAN	0.447	JEN	0.074	RIV	0.611	
PAR	0.941	13	BAY	0.148	MEL	0.213	ROK	0.743	
PAT	0.860	19	BAY	0.704	MEL	0.525			
RIT	0.990	10	BAY	0.723	MEL	0.475			
RIV	1.000	1	RIV	1.000					1
ROK	1.000	1	ROK	1.000					5
UCT	0.825	20	BAY	0.987	MEL	0.213			
VPK	1.000	1	VPK	1.000					1
WEB	0.872	17	BAY	1.021	MEL	0.272			

[a]#Ref: number of times this efficient DMU is a reference for an inefficient DMU.

the remaining inefficient districts, and the model does that, as indicated in the table.

We have omitted the sensitivity analysis for the DMUs under Model 2, as DEA-Solver does not calculate slacks for the CAT-O-C model.

4.3. Model 3

Model 3 was constructed exactly like Model 1, with the addition of one more input variable. The rich-and-poor categorical variable (3 = RICH) used in Model 2 to distinguish categories was used as an input variable to a CCR-O model as in Model 1. The result can be compared to both Models 1 and 2.

Table 7 shows the results for Model 3. This model found 10 efficient districts, one more than Model 2. That district is Ferguson-Florissant (FFL), also categorized as a poor district. The remaining nine efficient

Table 7. Model 3 (CCR-O with Additional Input Variable) Solutions of Efficiency Score and Reference Set.

DMU	Score	Rank	Reference Set (DMUs and Intensities)								#Ref[a]
AFT	0.976	12	BAY	0.793	HAN	0.302	MEL	0.056			
BAY	1.000	1	BAY	1.000							11
BRN	1.000	1	BRN	1.000							0
CLA	0.811	22	BAY	1.428	HAN	0.062	ROK	0.027			
FFL	1.000	1	FFL	1.000							0
HAN	1.000	1	HAN	1.000							6
HAZ	1.000	1	HAZ	1.000							1
JEN	1.000	1	JEN	1.000							1
KRK	0.862	19	BAY	0.952	MEL	0.341	ROK	0.013			
LAD	0.921	17	BAY	1.264	ROK	0.086					
LIN	0.956	13	BAY	0.833	MEL	0.266	ROK	0.070			
MEL	1.000	1	MEL	1.000							8
MRH	0.935	16	BAY	0.403	HAN	0.051	VPK	0.572			
NOR	0.841	20	HAN	0.734	HAZ	0.143	RIV	0.124			
PAR	0.941	15	BAY	0.148	MEL	0.213	ROK	0.743			
PAT	0.893	18	BAY	0.290	HAN	0.381	MEL	0.509	VPK	0.011	
RIT	0.994	11	BAY	0.492	MEL	0.508					
RIV	1.000	1	RIV	1.000							2
ROK	1.000	1	ROK	1.000							5
UCT	0.834	21	BAY	0.789	JEN	0.035	MEL	0.057	RIV	0.271	
VPK	1.000	1	VPK	1.000							3
WEB	0.950	14	BAY	0.197	HAN	0.378	MEL	0.311	VPK	0.302	

[a]#Ref: number of times this efficient DMU is a reference for an inefficient DMU.

districts are the same ones found by Model 2. Some have the same scores in both models; some are slightly different. Their reference sets however are different, reflecting the different methods for calculation.

Table 8 presents the sensitivity analysis for Model 3. Its interpretation is similar to that of Model 1.

4.4. Model Comparisons

As promised by the theory of the CAT methodology, Model 2 shows significantly different results from Model 1. Under Model 2, the four additional districts may be construed as performing significantly better than might have been perceived under Model 1, and they are performing as well

Table 8. Model 3 (CCR-O with Additional Input Variable) Sensitivity Analysis for Input/Output Variables.

DMU	Excess CUREXP, $	Excess TEACHSAL, $	Excess NUMTEACH	Excess 3 = RICH	Shortage ADA	Shortage AYP	Shortage ACT	Shortage GRAD
AFT	4,206,974	3,005	0.0	0.0	0.0	0.0	0.5	8.5
BAY	0	0	0.0	0.0	0.0	0.0	0.0	0.0
BRN	0	0	0.0	0.0	0.0	0.0	0.0	0.0
CLA	18,465,280	0	47.0	0.0	0.0	0.9	0.0	16.3
FFL	0	0	0.0	0.0	0.0	0.0	0.0	0.0
HAN	0	0	0.0	0.0	0.0	0.0	0.0	0.0
HAZ	0	0	0.0	0.0	0.0	0.0	0.0	0.0
JEN	0	0	0.0	0.0	0.0	0.0	0.0	0.0
KRK	11,656,190	0	0.0	0.4	0.0	1.3	0.0	9.5
LAD	16,849,487	0	63.2	0.2	0.0	1.2	0.0	16.4
LIN	7,237,307	0	0.0	0.6	0.0	0.1	0.0	6.4
MEL	0	0	0.0	0.0	0.0	0.0	0.0	0.0
MRH	2,354,560	0	0.0	0.0	240.2	1.6	0.8	0.0
NOR	9,757,526	9,535	0.0	0.0	0.0	4.1	0.0	8.9
PAR	29,724,871	0	0.0	0.0	0.0	1.3	0.0	4.9
PAT	17,934,663	0	0.0	0.0	94.4	0.0	0.0	8.1
RIT	6,029,118	8,223	0.0	0.0	0.0	4.0	1.0	2.2
RIV	0	0	0.0	0.0	0.0	0.0	0.0	0.0
ROK	0	0	0.0	0.0	0.0	0.0	0.0	0.0
UCT	7,087,818	0	0.0	0.0	0.0	2.3	0.3	0.0
VPK	0	0	0.0	0.0	0.0	0.0	0.0	0.0
WEB	5,092,873	0	0.0	0.0	416.7	0.0	0.0	3.4

as might be expected, given their handicap. The districts that are still inefficient require special management attention, whether or not they have a handicap.

A comparison of Model 2 to Model 3 requires slightly more interpretation. When comparing the scores of the inefficient districts, four of the inefficient districts have the same efficiency score under both models. But the remaining districts have significantly higher scores under Model 3 than under Model 2. A check of the correlation matrix shows the categorical variable (3 = RICH) is highly and positively correlated to three of the four output variables (AYP, ACT, and GRAD). Therefore, explicitly including this variable as an input has strengthened the CCR-O model.

A comparison of the correlation matrices for Model 3 and Model 1 confirms this conclusion. No input variable in Model 1 is as strongly correlated to those three outputs (AYP, ACT, and GRAD) as is 3 = RICH.

We therefore conclude that use of a categorical variable such as the one described here may be an alternative method for performing a categorical analysis in DEA. Additional tests using additional DMUs and additional variables are needed to confirm or deny this conclusion.

5. CONCLUDING REMARKS

Managers of public school districts face unprecedented challenges. Population migration, demographic shifts, political considerations, and funding variations all contribute to outcome uncertainty and stress to those who allocate educational resources. The continued existence of many public school districts is threatened in an unprecedented manner.

DEA has proven to be a valuable tool for analyzing and allocating educational resources. In this chapter, we have demonstrated its use in 22 school districts in Saint Louis County and identified a mix of efficient and inefficient districts.

We have also introduced a simple analytical tool for categorizing districts on the basis of their composite wealth and have demonstrated the utility of this tool in performing DEA analyses. It has been shown to have relevancy for both CCR-O and CAT variations of the DEA model.

With three DEA models, we found a different decision outcome when including the rich-and-poor (wealth) variable than when excluding it. Our baseline model (excluding the wealth variable) produced a less robust decision outcome than did either of the two alternatives, both of which included the wealth variable.

All of the three models presented in this chapter are easy to formulate and interpret. Data required by the models is readily available in the public domain and known to all school administrators in Missouri. This chapter has demonstrated that consideration of a district wealth indicator, while simple to use, can alter the decisions one would make in the absence of such a variable. On the basis of data availability and data modeling tools, the continued exclusion of a wealth variable in state funding formulas and accreditation models cannot be justified.

Although this study has demonstrated the benefit of a wealth variable, additional study is needed to refine the mix of environmental variables used to construct the variable. Various approaches such as principal component analysis (PCA) or OLS should be explored. The method should then be tested with a larger sample such as the entire Missouri schools data base. Additional studies are also needed to develop a model that incorporates a wealth variable with DEA projections onto the efficient frontier, as a means of improving equity when allocating funds to districts. Finally, additional studies are needed to develop an improved accreditation model that would incorporate a wealth variable.

NOTE

1. A 24th district encompasses the entire County geographically. It is organized to provide special education services to students in the other 23 districts and also serves as the County's Area Vocational-Technical School (AVTS).

ACKNOWLEDGMENT

We wish to thank Martha Ann Garrett, Special School District of Saint Louis County, for her assistance in gathering and interpreting the statistical data used for this study.

REFERENCES

Bessent, A. M., & Bessent, E. W. (1980). Determining the comparative efficiency of schools through data envelopment analysis. *Educational Administration Quarterly*, *16*(2), 57–75.

Bessent, A. M., & Bessent, E. W. (1981). *Productivity in community college programs: A technique for determining relative efficiency*. Dallas, TX: The Community College Productivity Center.

Charnes, A., Cooper, W. W., & Rhodes, E. L. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, *2*(6), 429–444.

Cooper, W. W., Seiford, L. M., & Tone, K. (2007). *Data envelopment analysis* (2nd ed.). New York: Springer.

Dunson, M. (2009). State budget shortfall forces hard choices for education, others, Joplin Globe, January 4. Available at http://www.joplinglobe.com/homepage/local_story_004000133.html/resources_printstory. Retrieved on May 30, 2009.

Kwak, N. K., & Kim, S. H. (2000). Data envelopment analysis: Concepts, applications, and perspectives. In: S. H. Dahiya (Ed.), *The current states of business disciplines* (Vol. 2, pp. 519–535). Rohtak, India: Spellbound Publications, Ltd.

Kwak, N. K., & Lee, C. W. (2009). Sustainability assessment of venture business firms using data envelopment analysis. In: K. D. Lawrence & G. Kleinman (Eds), *Applications of management science* (Vol. 13, pp. 247–260). Bingley, UK: Emerald Group Publishing Ltd.

Lieb, D. (2009). Missouri high court hears school funding challenge, *Columbia Missourian,* May 19. Available at http://www.columbiamissourian.com/stories/2009/05/19/missouri-high-court-hears-school-funding-challenge/. Retrieved on May 29, 2009.

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: U.S. Government Printing Office.

Primont, D. F., & Domazlicky, B. (2006). Student achievement and efficiency in Missouri schools and the no child left behind act. *Economics of Education Review*, *25*(1), 77–90.

Ruggiero, J., & Vitaliano, D. F. (1999). Assessing the efficiency of public schools using data envelopment analysis and frontier regression. *Contemporary Economic Policy*, *17*(3), 321–331.

Welker, J. L. (2006). *A study of the school funding formula created by SB 287 in Missouri*. Columbia, MO: University of Missouri-Columbia.

SUSTAINABILITY ASSESSMENT OF VENTURE BUSINESS FIRMS USING DATA ENVELOPMENT ANALYSIS

N. K. Kwak and Chang Won Lee

ABSTRACT

An appropriate assessment of sustainability in venture business is an important managerial and investment decision making. Data envelopment analysis (DEA) is utilized for sustainability assessment for venture business firms' performance. Venture business firms are primary decision-making units (DMUs). Required information for this study is collected from Korea Listed Companies Association (KLCA) database. The proposed DEA model incorporates multiple inputs and outputs to assess the relative operational efficiency of the DMUs, identifying the best performance group among the peer venture business firms. The proposed model provides decision-makers with more accurate information for strategic insights to make better investment decisions in the competitive business environment.

Financial Modeling Applications and Data Envelopment Applications
Applications of Management Science, Volume 13, 247–260

ISSN: 0276-8976/doi:10.1108/S0276-8976(2009)0000013015

1. INTRODUCTION

An effective sustainability assessment of venture firms is an important managerial issue in terms of business decision-making for operational strategy. The problem of sustainability assessment in venture firm settings has become a significant and strategic matter for a firm's sustainability measurement. The strategies relevant to venture business management are interactive to operational efficiency with sustainability. These issues are complex and conflicting decision-making matters (Beasley, 2003). Most quantitative modeling techniques that are available to assess venture sustainability are classified as heuristic, multicriteria decision-making (MCDM) models, or simulation. However, few studies have explored the venture sustainability evaluation issues in terms of operational excellence using data envelopment analysis (DEA).

DEA is a nonparametric mathematical programming technique (Adler & Golany, 2001; Wang, Gopal, & Zionts, 1997). It was developed by Charnes, Cooper, and Rhodes (1978) as a tool for empirically evaluating the relative efficiency of productive organizations, called decision-making units (DMUs). There are also parametric approaches used for the estimation of productive efficiency (Lovell & Schmidt, 1988) and a hybrid of combining the relative strengths of parametric and nonparametric approaches (Tofallis, 2001).

The purpose of this study is to evaluate the performance efficiency of venture system of 30 venture DMUs. Specifically, this study focuses on developing and analyzing a DEA mode1: (1) to determine the relative efficiency of venture system performance in each member venture business firms, (2) to estimate the amounts of identified inefficiencies, and (3) to compare performance between the efficient venture firms and the inefficient venture firms to provide management with a strategic insight for better planning and controlling operations of venture system. This DEA model is able to enhance a decision-making process and planning policy for performance measurements in similar settings. The DEA model in this study positions the performance assessment planning situation in a venture system to respond to better customer satisfaction, appropriate efficiency measurement, and effective resource allocation, while strengthening ongoing planning strategies to meet defined requirements of the venture system.

The chapter is organized in the following manner. Section 2 deals with a review of DEA literature. Section 3 presents a problem background with data modeling. Section 4 illustrates model development dealing with input–output variables and model formulation. Section 5 provides the model analysis and discussion, along with concluding remarks.

2. DATA ENVELOPMENT ANALYSIS

DEA is an application of a mathematical programming technique often applied to MCDM problems (Lovell & Schmidt, 1988; Seiford & Thrall, 1990). It has no assumptions about the form of production function of the productive organization. It estimates the best practice frontier from the observed input–outputs of DMUs. In the DEA analysis, the organization's operating units are conceived as consuming a set of inputs to produce a set of outputs. DEA has been established as one of the accepted methodologies of operations research/management science in the past few decades as a relative evaluation of operational performance of the productive and service systems. DEA actually encompasses a variety of alternative approaches to evaluating the organization's performance. The organization's efficiency is often measured by the ratio of output to input. When there is more than one input and output, a set of weights for aggregating these inputs (outputs) into a single virtual input (output) can be used to obtain a ratio measure of efficiency.

Building on the ideas of measuring the productive efficiency of Farrell (1957) and Charnes et al. (1978) developed DEA as a multiplicative model for evaluating the relative efficiency of DMUs with the same goals and objectives. It is referred to as the CCR model, named after its originators (Charnes, Cooper, and Rhodes).

The CCR model assumed constant returns-to-scale in a performance analysis, meaning that a change in the amount of all inputs leads to the same change in the amount of all outputs. Banker, Charnes, and Cooper (1984) introduced the modified DEA model incorporating variable returns-to-scale for estimating technical and scale inefficiencies in DEA studies. It is named the BCC model. The CCR model and the BCC model are the two models used most frequently in DEA research.

The generalized CCR model is presented below (Eq. (1)):

$$\begin{aligned}
\text{Maximize}: \quad & E_k = \frac{\sum_{r=1}^{s} u_r y_{rk}}{\sum_{i=1}^{m} v_i x_{ik}} \\
\text{Subject to}: \quad & \frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \leq 1; \quad \text{for } j = 1, 2, \ldots, n \\
& u_r \geq \varepsilon; \quad \text{for } r = 1, 2, \ldots, s \\
& v_i \geq \varepsilon \ \text{ for } i = 1, 2, \ldots, m
\end{aligned} \tag{1}$$

where

y_{rj} = observed amount of rth output for the jth DMU
x_{ij} = observed amount of ith input for the jth DMU
u_r = weight attached to the rth output
v_i = weight attached to the ith input
n = number of DMUs
s = number of outputs
m = number of inputs
ε = a non-Archimedean small positive number

The decision variables in this model are the weights u_r ($r = 1,2, \ldots, s$) and v_i ($i = 1,2, \ldots, m$). The objective function is the efficiency of DMU_k to be determined through the value of weights. The efficiency of DMU_k is the ratio of the virtual output $\sum_{r=1}^{s} u_r y_{rk}$ of DMU_k to its virtual input $\sum_{i=1}^{m} v_i x_{ik}$. The weights u_r and v_i are allowed to take any strictly positive values to maximize the efficiency of DMU_k, given that the ratio for any DMU under consideration does not exceed 1. To obtain the efficiency scores of all DMUs, it is required to solve n problems, each of which differs only in the coefficients in its objective function.

In Eq. (1), the ratio of virtual output-to-input is not linear, and thus, Eq. (1) is a nonlinear programming problem. It is converted into a linear programming problem by adding the constraint that the virtual inputs and outputs are one (i.e, $\sum_{i=1}^{m} v_i x_{ik} = 1$) and an assumption that all inputs and outputs are positive, using a transformation theory developed by Charnes and Cooper (1962). It is referred to as the multiplier form of the CCR model and is presented in Eq. (2) as follows:

$$
\begin{aligned}
\text{Maximize}: \quad & E_k = \sum_{r=1}^{s} u_r y_{rk} \\
\text{Subject to}: \quad & \sum_{i=1}^{m} v_i x_{ik} = 1 \\
& \sum_{i=1}^{m} u_r y_{ij} - \sum_{i=1}^{m} v_i x_{ij} \leq 0; \quad \text{for } j = 1, 2, \ldots, n \\
& u_r \geq \varepsilon; \quad \text{for } r = 1, 2, \ldots, s \\
& v_i \geq \varepsilon; \quad \text{for } i = 1, 2, \ldots, m
\end{aligned}
\tag{2}
$$

In Eqs. (1) and (2), the efficiency E_k is always less than or equal to 1 because the objective function is also part of a constraint. The DMU_k is efficient relative to the other DMUs if $E_k = 1$, and it is inefficient if $E_k < 1$.

The BCC model is slightly different from the CCR model (Eq. (2)) by the addition of a variable u_k as shown in Eq. (3) below:

$$\begin{aligned}
\text{Maximize :}\quad & E_k = \sum_{r=1}^{s} u_r y_{rk} - u_k \\
\text{Subject to :}\quad & \sum_{i=1}^{m} v_i x_{ik} = 1 \\
& -\sum_{i=1}^{m} v_i x_{ij} + \sum_{r=1}^{s} u_r y_{rj} - u_k \leq 0;\quad \text{for } j = 1,2,\ldots,n \\
& u_r \geq \varepsilon;\quad \text{for } r = 1,2,\ldots,s \\
& v_i \geq \varepsilon;\quad \text{for } i = 1,2,\ldots,m
\end{aligned} \tag{3}$$

The variable u_k serves as a convexity constraint, allowing the efficiency frontier to envelop the observed data more tightly than the CCR model (Eq. (2)).

Since the inception of DEA, it has been applied to a wide variety of organizations. There have been several comprehensive DEA bibliographies such as Charnes, Cooper, Lewin, and Seiford (1994), Seiford (1996), Serafoglou (1998), and Gattoufi, Oral, and Reisman (2004). For example, Seiford listed over 700 DEA applications. Kwak and Kim (2000) provided a summary of some DEA applications not covered by Charnes et al. and Seiford. Recently, DEA has been utilized for analyzing relative operational efficiency of various service areas such as airport (Sarkis, 2000), information (Vargas, Hernandez, & Bruque, 2003), railway transportation (Kwak, Choi, & Kim, 2004), hotel operations (Botti, Briec, & Cliquet, 2009), and sports (Cooper, Ruiz, & Sirvent, 2009), to name a few.

3. DATA BACKGROUND AND MODEL DEVELOPMENT

An appropriate assessment of sustainability in venture business operations is an important managerial and investment decision-making process. Strategic objectives in venture business operations are to enhance organizational efficiency by streamlining key decision-making processes and to develop an integrated venture system. In this study, DEA is applied to assess the sustainability of venture business firms' operational performance. In the DEA model development, the venture business operation is considered as a production process, utilizing four input variables (i.e., total amounts of assets, employees, R&D expenditures, and administrative-operating expenses) and four output variables (i.e., growth rate, profitability, stability, and net sales). The necessary data in this study is obtained from the Korea Listed Companies Association (KLCA) publications. Although the 30 selected firms are derived from diverse industrial sectors, they have been classified as the most outstanding venture firms among the entire venture firms listed on the Korea Securities Dealers Automated Quotation (KOSDAQ). Thus, these firms will provide homogeneity in terms of business performance aspects. (See Appendix for the 30 selected venture business firms.)

Banker et al. (1984) and Boussofiane, Dyson, and Thanassoulis (1991) suggested the rules on the number of input and output variables to be used in the DEA applications. To select appropriate input and output variables, a panel of group decision-makers for venture firms is involved in the model formulation planning. After the generation of necessary data templates related to the model formulation planning, the panel reviews the data set and provides an aggregated opinion of the validation for the collected data.

The four input variables are: total asset (ASET) represented by the total amount of assets ($000) that the venture firm has for its operations; employee (EMPL) shown by the total number of employees at the term-end in a given year; research and development (RNDE) including the total R&D expenses ($000) in a given year; and administrative expense (ADME) including the total operating and administrative expenses ($000) in a given year. Four output variables are growth (GROW) represented by a value added per employee growth rate; profitability (PROF) on earnings per share; stability (STAB), on operating capital share; and net sales (NSAL) amount($000). A summary of these input and output variables is presented in Table 1.

Table 1. Input and Output Variables.

Variables	Descriptions
Inputs (venture firm resources)	• Total asset (ASET): total amount of assets ($000) • Total employees (EMPL): total number of employees • R&D expenditures (RNDE): total R&D expenses ($000) • Administrative expense (ADME): total operating and administrative cost ($000)
Outputs (venture firm outcomes)	• Growth (GROW): value added per employee growth rate • Profitability (PROF): earnings per share • Stability (STAB): operating capital per share • Net sales (NSAL): net sales amount ($000)

Table 2 presents the necessary data for input and output variables with respect to each DMU in year 2008. DEA architecture is designed to be a four-input and four-output variable format. A technology structure of DEA is selected with a convex envelopment and constant returns-to-scale

Table 2. Input and Output Variable Characteristics for DMUs.

Values	Input				Output			
	ASET ($000)	EMPL (000)	RNDE ($000)	ADME ($000)	GROW (%)	PROF ($000)	STAB (%)	NSAL ($000)
CJ Home	889	705	1,653	435	−13.39	2,889	39.84	519
CJ Int	229	337	0	25	53.62	1,134	16.48	160
Credu	79	214	70	11	0	1,959	15.13	63
Daum	221	681	0	163	−4.93	1,227	61.6	215
Dongwha	311	85	0	12	−35.1	17	13.87	18
Eugene	860	865	133	62	−62.74	392	64.02	441
Forhuman	105	44	133	2	51.11	651	1.71	9
GS Home	581	863	937	459	−5.36	7,462	61.22	593
Hana	184	1,502	0	163	−10.21	2,352	84.38	199
Humax	645	577	41,009	94	−33.5	606	46.85	576
Hyunjin	261	247	299	8	−11.87	1,676	61.76	222
JVM	99	217	246	9	−8.11	2,203	25.55	46
Jusung	274	416	36,578	62	68.98	1,248	42.08	212
Kiwoom	1,308	352	0	98	17.88	5,754	1.58	493
MegaStu	206	515	1,633	44	28.72	7,290	26.2	163
Mode	97	841	0	80	−2.62	1,360	62.5	94
POSDATA	288	1,350	615	40	−1.69	61	51.3	365
PyeongSan	403	279	0	18	−10.43	1,600	77.63	255
SFA Eng	377	549	0	18	−20.02	4,615	76.67	307
SK Brod	2,671	1,615	0	1,787	6.08	31	28.52	1,868
SK Comm	307	1,133	1,098	174	16.28	−992	36.69	197
SSCP	392	455	886	27	0.09	889	45.82	184
Seoul Se	225	984	7,886	43	19.55	356	21.91	250
Sodiff	277	249	2,758	13	57.79	1,821	60.16	98
Ssangyong	1,164	1,220	1,348	110	−13.38	1,190	129.4	1,336
SungKwang	241	354	0	16	112.53	1,791	72.01	258
TK	229	445	1,087	21	32.39	2,271	35.24	287
Taewoong	265	256	1,115	15	15.39	3,101	70.7	358
Techno	194	373	6,806	16	6.72	1,797	25.66	167
Unison	315	348	2,754	12	−19.70	143	79.27	61
Mean	456	602	3,760	134	8	1,896	48	333
MIN	79	44	0	2	−63	−992	2	9
MAX	2,671	1,615	41,009	1,787	113	7,462	129	1,868

(CCR model) because of ease of measuring and a more meaningful interpretation of the results. The relevant results are displayed as the virtual inputs and outputs, meaning that appropriate weights are multiplied by the input and output values. An efficiency measure quantifies a distance to the efficient frontier of the technology. Thus, it quantifies the input reduction that is necessary to become efficient holding the output constant.

To measure the efficiency, a radial distance algorithm is used. This measure indicates necessary improvements when all relevant factors are improved by the same factor in the same proportion. It is referred to as a Debreu–Farrell measure or radial part of the CCR/BCC models (Chakravarty, 1992). This algorithm has fairly good price interpretations (cost reduction/revenue increase). A software system, *EMS* (Scheel, 2000), was utilized to conduct this study.

4. MODEL ANALYSIS AND DISCUSSION

Adequate assessments of venture firms' operational efficiency using DEA have received little attention in the business community, although their strategic importance in both the public and the private sectors has emerged. It is necessary to assess the relative operational performance efficiency of the individual DMU. The results of this DEA model are based on the four-input and four-output data with a constant return-to-scale, radial distance, maximization orientation, and weights (shadow prices) restriction with a restriction matrix. This requires a performance measurement system that can operate at several different DMU levels. For an efficient DMU, the reference DMUs with corresponding intensities are presented. For relative efficiency measurements of inefficient DMUs, the numbers of efficient DMUs selected as a benchmark are provided. Slacks are provided for a radial distance measure.

In Table 3, score 1 means that a venture firm's business system is operating efficiently. A total of 17 venture firms (Credu, Forhuman, Hana, Hyunjin, JVM, Kiwoom, Megastu, Mode, POSDATA, PyeongSan, SFA Eng, SK Brod, Sodiff, Ssangyong, SungKwang, Taewoong, and Unison) are measured as efficient when compared to the remaining 13 venture firms. Thus, these 17 venture firms use less input resources than the remaining 13 venture firms to produce a given level of business performance.

The fourth column of Table 3 illustrates some comparison of the reference venture firms for inefficient venture firms and the relative importance of reference venture firms (in parentheses). The reference comparison helps to

Table 3. Solutions of Efficiency Score and Intensity Levels for Reference.

No	DMU	Score	Benchmark				Ref DMU[a]
1	CJ Home	0.59	14 (0.01)	20 (0.10)	26 (0.09)	28 (0.87)	
2	CJ Int	0.66	9 (0.00)	19 (0.01)	26 (0.61)		
3	Credu	1.00					1
4	Daum	0.91	9 (0.26)	26 (0.63)			
5	Dongwha	0.59	18 (0.18)				
6	Eugene	0.60	18 (0.64)	20 (0.00)	25 (0.06)	26 (0.74)	
7	Forhuman	1.00					0
8	GS Home	0.98	3 (1.13)	19 (0.47)	26 (0.43)	28 (0.75)	
9	Hana	1.00					2
10	Humax	0.71	14 (0.04)	28 (1.56)			
11	Hyunjin	1.00					0
12	JVM	1.00					0
13	Jusung	0.68	26 (0.59)	28 (0.17)			
14	Kiwoom	1.00					2
15	MegaStu	1.00					1
16	Mode	1.00					0
17	POSDATA	1.00					1
18	PyeongSan	1.00					3
19	SFA Eng	1.00					2
20	SK Brod	1.00					2
21	SK Comm	0.49	17 (0.02)	26 (0.08)	28 (0.47)		
22	SSCP	0.42	18 (0.02)	26 (0.28)	28 (0.34)		
23	Seoul Se	0.84	26 (0.09)	28 (0.64)			
24	Sodiff	1.00					0
25	Ssangyong	1.00					1
26	SungKwang	1.00					10
27	TK	0.97	26 (0.20)	28 (0.66)			
28	Taewoong	1.00					9
29	Techno	0.66	15 (0.06)	28 (0.44)			
30	Unison	1.00					0

[a]Ref DMU: the number of appearances as a reference DMU.

highlight a DMU's weakness. It shows the relative performance of the unit compared with its closest peers (reference set). For example, the reference venture firms of CJ Home (CJ Home Shopping) are 14-Kiwoom (Kiwoom Securities), 20-SK Brod (SK Broadband), 26-SungKwang (Sung Kwang Bend Co., Ltd.), and 28-Taewoong (Taewoong Co., Ltd.).

It means these four venture firms are utilized to measure the efficiency score of CJ Home (DMU No. 1). To compare among the efficient DMUs, the number of appearances as reference DMUs is usually used.

SungKwang (DMU No. 26) is used ten times and Taewoong (DMU No. 28) nine times for benchmarking. SungKwang (DMU No. 26) is most frequently used to evaluate the performance efficiency of DMUs; Taewoong and PyungSan are used nine and three times, respectively, to evaluate the DMU efficiency.

Table 4 presents the results of sensitivity analysis for four-input and four-output variables. Input slack indicates an input variable having an idle

Table 4. Sensitivity Analysis for Input/Output Variables.

DMU	Score	SR1	SR2	SR3	SR4	SO1	SO2	SO3	SO4
1	0.59	0	0	0	65590	37	0	31	0
2	0.66	0	0	0	6290	14	0	28	0
3	1.00								
4	0.91	0	0	0	94406	73	516	6	0
5	0.59	110408	0	0	3586	33	269	0	27
6	0.60	0	0	0	0	138	2028	47	0
7	1.00								
8	0.98	0	0	0	408857	55	0	75	0
9	1.00								
10	0.71	0	0	27558	40678	58	4442	63	0
11	1.00								
12	1.00								
13	0.68	0	31	24658	29780	0	326	12	0
14	1.00								
15	1.00								
16	1.00								
17	1.00								
18	1.00								
19	1.00								
20	1.00								
21	0.49	0	377	0	75976	0	2597	4	0
22	0.42	0	0	0	1333	37	694	0	15
23	0.84	0	635	5928	25659	0	1773	29	0
24	1.00								
25	1.00								
26	1.00								
27	0.97	0	193	320	7494	0	128	26	0
28	1.00								
29	0.66	0	103	3911	1798	2	0	7	0
30	1.00								

Note: SR1: Slack in ASET; SR2: Slack in EMPL; SR3: Slack in RNDE; SR4: Slack in ADME.SO1: Slack in GROW; SO2: Slack in PROF; SO3: Slack in STAB; SO4: Slack in NSAL.

Table 5. Group Statistics of Two Ventures Firms.

Test Variable	Groups	N	Mean	SD	SE
Stock price	Efficiency	13	7079.23	11736.512	3255.123
	Inefficiency	17	33205.29	37535.038	9103.584

Table 6. Independent Samples *t*-Test of Two Ventures Firms.

		Levene's Test[a]		*t*-Test for Equality of Means				
		F	Sig.	t	df	Sig. (Two-tailed)	Mean Difference	Standard Error Difference
Stock price	Equal variances assumed	4.607	.041	−2.412	28	.023	−26126.063	10830.498
	Equal variances not assumed			−2.702	19.919	.014	−26126.063	9668.043

[a]Levene's test for equality of variances.

capacity in ASET, EMPL, RNDE, or ADME. DMU No. 5 (Dongwha) has the highest slack in ASET; DMU No. 23 (Seoul Se) in EMPL; DMU No.10 (Humax) and DMU No. 13 (Jusung) in RNDE; and DMU No. 8 (GS Home) in ADME. The potential improvements indicate the target input and output levels needed for a DMU to become fully efficient.

Among the four output variables, DMU No. 6 (Eugene) has the most potential improvement in GROW; DMU No. 10 (Humax) in PROF; DMU No. 8 (GS Home) in STAB; and DMU No. 5 (Dongwha) in NSAL.

Tables 5 and 6 exhibit the descriptive statistics and *t*-test results for difference between the efficient group and the inefficient group. Efficient group means companies with efficiency score 1 (that is, $E_k = 1$), and inefficient group means companies with an efficiency score of less than 1 (that is, $E_k < 1$). To compare the two groups' performance in terms of firm's efficiency, a *t*-test has been performed with stock price variation of the firms. The difference between the closing prices in January and December of 2008 has been used as a test variable.

The *t*-test statistics in Table 6 indicate that the two venture groups (efficient and inefficient) have a statistically significant difference over the stock price variation in year 2008. Thus, this may suggest that efficient firms have better stock price performance, resulting in better sustainability in their business operations.

5. CONCLUDING REMARKS

This study presents an application of DEA to measure venture firms' operational efficiency and to predict the best practices in venture firms' sustainability performance. A DEA model was developed and analyzed to provide decision-makers with strategic insights for better management of venture system. The DEA CCR model was applied to assess venture firms' sustainability in business settings. The model incorporated multiple input and output variables to perform an appropriate analysis to assess the relative operational efficiency of the 30 selected DMUs. Required business profiles for this study were collected from KLCA database of year 2008.

Even though DEA techniques have been applied to various areas (e.g., business, health services research, public sector, and transportation), this study makes a substantial contribution to the venture firms' sustainability area. The DEA model developed in this study provides an opportunity for managing sustainability measurement in the planning process for venture firms. The model can be easily extended for application to any other entity or organization that faces a similar sustainability measurement in its planning environment.

ACKNOWLEDGMENT

This study is partly supported by a research grant of Jinju National University Graduate School of Entrepreneurial Study of 2007.

REFERENCES

Adler, N., & Golany, B. (2001). Including principal component weights to improve discrimination in data envelopment analysis. *Journal of the Operational Research Society*, *53*(9), 985–991.

Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, *30*, 1078–1092.

Beasley, J. E. (2003). Allocating fixed costs and resources via data envelopment analysis. *European Journal of Operational Research*, *147*(1), 198–216.

Botti, L., Briec, W., & Cliquet, G. (2009). Plural forms versus franchise and company-owned systems: A DEA approach of hotel chain performance. *Omega*, *37*(3), 566–578.

Boussofiane, A., Dyson, R. G., & Thanassoulis, E. (1991). Applied data envelopment analysis. *European Journal of Operational Research*, *52*(1), 1–16.

Chakravarty, S. R. (1992). Efficiency and concentration. *Journal of Productivity Analysis*, *3*(3), 249–255.

Charnes, A., & Cooper, W. W. (1962). Programming with linear fractional functionals. *Naval Research Logistics Quarterly*, *9*(3/4), 181–185.

Charnes, A., Cooper, W. W., Lewin, A. Y., & Seiford, L. (1994). *Data envelopment analysis: Theory, methodology, and applications*. Norwell, MA: Kluwer Academic Publishing.

Charnes, A., Cooper, W. W., & Rhodes, E. L. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, *2*(6), 429–444.

Cooper, W. W., Ruiz, J. L., & Sirvent, I. (2009). Selecting non-zero weights to evaluate effectiveness of basketball players with DEA. *European Journal of Operational Research*, *195*(2), 563–574.

Farrell, J. M. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society*, *120*, 253–281.

Gattoufi, S., Oral, M., & Reisman, A. (2004). A taxonomy for data envelopment analysis. *Socio-Economic Planning Sciences*, *38*(2–3), 115–232.

Kwak, N. K., Choi, T. S., & Kim, S. (2004). Evaluating the relative efficiency of railway operations using data envelopment analysis. In: K. D. Lawrence (Ed.), *Mathematical programming: Applications of management science* (Vol. 11, pp. 77–88). The Netherlands: Elsevier Amsterdam.

Kwak, N. K., & Kim, S. H. (2000). Data envelopment analysis: Concepts, applications, and perspectives. In: S. H. Dahiya (Ed.), *The current states of business disciplines* (Vol. 2, pp. 519–535). Rohtak, India: Spellbound Publications Ltd.

Lovell, C. A. L., & Schmidt, P. (1988). A comparison of alternative approaches to the measurement of productive efficiency. In: A. Dogramaci & R. Färe (Eds), *Applications of modern production theory: efficiency and productivity* (pp. 3–32). Boston: Kluwer.

Sarkis, J. (2000). An analysis of the operational efficiency of major airports in the United States. *Journal of Operations Management*, *18*(3), 335–351.

Scheel, H. (2000). EMS: Efficiency measurement system. Available at http://www.wiso.uni-dortmund.de/lsfg//or/scheel/ems/

Seiford, L. M. (1996). Data envelopment analysis: The evolution of the state of art (1978–1995). *Journal of Productivity Analysis*, *7*(2–3), 99–137.

Seiford, L. M., & Thrall, R. M. (1990). Recent developments in DEA: The mathematical programming approach to frontier analysis. *Journal of Econometrics*, *46*, 7–38.

Serafoglou, N. (1998). The most influential DEA publications: A comment on Seiford. *Journal of Productivity Analysis*, *9*(3), 279–281.

Tofallis, C. (2001). Combining two approaches to efficiency assessment. *Journal of the Operational Research Society*, *52*(11), 1225–1231.

Vargas, A., Hernandez, M. J., & Bruque, S. (2003). Determinants of information technology competitive value, evidence from a Western European industry. *Journal of High Technology Management Research*, *14*(2), 245–268.

Wang, C. H., Gopal, R. D., & Zionts, S. (1997). Use of data envelopment analysis in assessing information technology impact on firm performance. *Annals of Operations Research*, *73*, 191–214.

APPENDIX. FULL NAMES OF DECISION-MAKING UNITS (DMUS)

DMU Acronym	Full name
CJ Home	CJ Home Shopping Co., Ltd.
CJ Int	CJ Internet Corp.
Credu	Credu Corporation
Daum	Daum Communications Corp.
Dongwha	Dongwha Holdings Co., Ltd.
Eugene	Eugene Corporation
Forhuman	Forhuman Co., Ltd.
GS Home	GS Home Shopping Inc.
Hana	Hana Tour Service Inc.
Humax	Humax Co., Ltd.
Hyunjin	Hyunjin Materials Co., Ltd.
JVM	JVM Co., Ltd.
Jusung	Jusung Engineering Co., Ltd.
Kiwoom	Kiwoom Securities Co., Ltd.
MegaStu	Megastudy Co., Ltd.
Mode	Modetournetwork.Inc.
POSDATA	POSDATA Company, Ltd.
PyeongSan	Pyeong San Co., Ltd.
SFA Eng	SFA Engineering Corp.
SK Brod	SK Broadband Co., Ltd.
SK Comm	SK Communications Co., Ltd.
SSCP	SSCP Co., Ltd.
Seoul Se	Seoul Semiconductor Co., Ltd.
Sodiff	Sodiff Advanced Materials Co., Ltd.
Ssangyong	Ssangyong Engineering & Construction Co., Ltd.
SungKwang	Sung Kwang Bend Co., Ltd.
TK	TK Corporation
Taewoong	Taewoong Co., Ltd.
Techno	Techno Semichem Co., Ltd.
Unison	Unison Co., Ltd.

DETERMINING THE RELATIVE EFFICIENCY OF GYNECOLOGICAL DEPARTMENTS USING DEA

Reuven R. Levary and Cesse Ip

ABSTRACT

Data envelopment analysis (DEA) is used to determine the relative efficiency of the top-ranked gynecology departments in the United States as designated by the U.S. News & World Report ranking. DEA is a linear programming base procedure used to determine the relative efficiency of operating units that have similar characteristics. Efficiency scores are calculated by comparing two different input sets to the performance of each gynecological department. Ranking based on DEA more completely and accurately represents gynecological departments. Further, DEA makes it possible to fairly compare specific departments. The new ranking coupled with the efficiency score accrued by each hospital will motivate and guide hospital administrators to improve the performance of hospital gynecology departments by better utilizing expensive resources.

Financial Modeling Applications and Data Envelopment Applications
Applications of Management Science, Volume 13, 261–273

ISSN: 0276-8976/doi:10.1108/S0276-8976(2009)0000013016

1. INTRODUCTION

Hospital rankings are often determined based on the number of beds, physician-to-nurse ratio, physician-to-patient ratio, nurse-to-patient ratio as well as on the number and type of prestigious services offered. Rankings that are based on the characteristics of input resources do not necessarily provide a good indication of how such resources affect measures of hospital performance.

In this study, we will examine different sets of resources, or inputs, and measure each hospital's efficient use of these resources to produce measurable outputs. A new ranking based on efficiency scores will be created. The new ranking and efficiency score associated with each hospital will assist hospital administrators in identifying those resources that can be handled more effectively to improve hospital performance. Improving the impact of input resources on hospital performance may affect both the quality of service provided to patients and the attractiveness of hospitals.

We chose to evaluate a specific type of hospital department instead of evaluating hospitals as a whole because some hospitals are well known for certain areas of medicine but might not excel in some other areas. Additionally, data envelopment analysis (DEA) has been used to rank hospitals as a whole (Hsu, Pi-Fang, & Hui-Chen, 2007; Chilinqerian & Sherman, 2004; Cruise & Nyhan, 2000) but to date has not been used to rank gynecological departments.

In this study, we will evaluate those gynecological departments in the United States that have been designated by the 2007 U.S. News & World Report ranking. Gynecology departments oversee births, fertility, and illnesses like ovarian and breast cancer. One out of eight women is affected by breast cancer in her lifetime. Globally, every year over one million women are diagnosed with it (Center For Disease Control, 2008). Additionally, there are over four million births a year in the United States. For these reasons, a large number of women make frequent visits to gynecology departments. Thus, it is advantageous to evaluate and rank gynecological departments according to the efficiency of converting resources into a measure of output performance.

2. USING DEA TO DETERMINE RELATIVE EFFICIENCY

Using DEA, the relative efficiency of decision-making units (DMUs) that use multiple inputs to produce multiple outputs may be calculated. The relative efficiency of a DMU is calculated using a ratio definition of efficiency (Charnes, Cooper, & Rhodes, 1978). This ratio generalizes the single output to single input definition to multiple outputs and inputs without the use of pre-assigned weights. The weights used for each DMU are those that maximize the ratio between the weighted output and weighted input. These weights are determined in such a way that no method of aggregating the inputs and outputs, such as value or market price, is necessary.

DEA is an analytical procedure developed by Charnes et al. (1978) for measuring the relative efficiency of DMUs that perform the same type of functions and have identical goals and objectives. DMUs include departments, sections, branches, and divisions of organizations belonging to the same business sector. If the relative efficiency of a set of DMUs performing the same type of function is to be evaluated, the DMUs must use the same type of input to produce the same type of output. Each DMU in a given set can then be ranked according to how efficiently it utilizes its inputs to produce its outputs.

When the combined number of inputs and outputs approaches the total number of DMUs in a set, however, DEA may be problematic. Under such circumstances, one must be very cautious in interpreting efficiency scores (Charnes, Cooper, Golany, Seifoed, & Stutz, 1985a).

Numerous refinements of DEA now enhance its analytical effectiveness. The "window analysis" concept (Charnes & Cooper, 1985) was incorporated into DEA to enable it to trace the performance of each DMU over time. Tracing performance over time is done by evaluating the DMUs at different time periods. As "window analysis" requires that a DMU be defined for each time period used in the analysis, it substantially increases the volume of calculations. Thanassoulis and Dyson (1992) developed several DEA-based models that can be used to estimate alternative input–output target levels and are helpful in rendering relatively inefficient organizational units efficient.

The DEA model applied in this study was developed by Banker, Charnes, and Cooper (1984) and has been used in many applications (e.g., Seiford, 1996; Colbert, Levary, & Shaner, 2000). Chang and Guh (1991) pointed out some problems of using this model. A comprehensive bibliography of DEA is given in Seiford (1996). The efficiency measure for each DMU ranges from 0 to 1. A DMU with an efficiency value of 1 is considered most efficient. An efficiency value smaller than 1 indicates the degree of relative efficiency. One possible explanation of a DMU's inefficiency is that some of its inputs are not utilized fully. Efficient DMUs achieve greater output per unit of input when compared with inefficient DMUs. By identifying unutilized resources, DEA can provide a firm's management with some information regarding causes of inefficiency.

In order to formulate the DEA model, let us assume that n hospitals are to be evaluated based on m inputs and s outputs. Let y_{rj} be a known level of the rth output of hospital j ($r = 1, 2, \ldots, s; j = 1, 2, \ldots, n$) and x_{ij} be a known level of the ith input to hospital j ($i = 1, 2, \ldots, m$). Each hospital is assigned a weight w_j ($j = 1, 2, \ldots, n$) for its input and output. A hypothetical composite hospital can then be defined using weighted inputs and outputs of the hospital being evaluated. The weights w_j are the model decision variables. The efficiency of hospital k relative to the composite hospital can be determined by solving the following linear programming problem:

$$\text{Min } h_k \tag{1}$$

$$\text{Subject to : } \sum_{j=1}^{n} w_j = 1 \tag{2}$$

$$\sum_{j=1}^{n} w_j y_{rj} \geq y_{rk}, \qquad r = 1, 2, \ldots, s \tag{3}$$

$$\sum_{j=1}^{n} w_j x_{ij} \leq x_{ik} h_k, \qquad i = 1, 2, \ldots, m \tag{4}$$

$$h_k, w_j (j = 1, 2, \ldots, n) \geq 0 \tag{5}$$

where h_k is the relative efficiency of hospital k.

Minimizing the relative efficiency of hospital k is equivalent to minimizing the inputs of the composite hospital program. Constraint (2) ensures that the sum of the weights is equal to 1. Constraint (3) ensures that each output level of the composite hospital is at least as high as the output level of hospital k. Constraint (4) ensures that each input level of the composite hospital is at most as high as its input capacity. A comprehensive coverage of DEA is provided by Cooper, Seiford, and Ton (2006).

3. OTHER METHODS OF DETERMINING EFFICIENCY

DEA is not the only method that can be used to determine efficiency (Colbert et al., 2000). Ratio analysis is another common method. While the

DEA is a ratio model (Charnes et al., 1978), we refer here to a ratio analysis method that is not a DEA-based method. Using ratio analysis, a ratio comparing outputs to inputs is computed. A simple ratio compares one measure of input to one measure of output.

Multiple inputs and outputs may be incorporated into ratio analysis by calculating multiple ratios. However, this makes it difficult to determine overall efficiency. A measure of overall efficiency can be computed by aggregating all inputs and outputs. This requires assigning a weight to each input and output. While such weights may be determined according to the value or market price of each input and output, this information is not always available. When the market value of each input and output is missing, one may consider using the Cook and Kress (1990) approach to determine the set of weights. Cook and Kress developed a methodology for aggregating preference ranking and applied their approach to aggregate votes in a preferential election. Their model determines for each candidate i the best set of weight w_j that should be applied to the jth place standing v_{ij}.

Multiple regression is another method for determining efficiency. Using multiple regression, output level is modeled as a function of various input levels. Operating units that are relatively efficient lie above the modeled relationship and have positive residual. Operating units that are relatively inefficient lie below the modeled relationship and have negative residuals. This method has several drawbacks. First, because single-equation multiple regression can model only one output level, a single output measure must be determined or all outputs must be artificially combined into a single indicator. Multiple-equation regression models can be used when an operating unit has multiple outputs. Like multiple ratios, however, this method does not produce an overall measure of efficiency. Multiple residuals provide different measures of the operating unit's efficiency.

Another drawback of regression analysis is that it compares efficiency with average performance rather than with the best performance. Additionally, "regression analysis requires the parametric specification of a production function, that is, an equation detailing how inputs are combined to produce outputs" (Sexton, 1986, p. 9). This is often difficult because such a function may be unknown for the industry in question.

Several studies combined DEA with regression analysis to evaluate operating units that have multiple inputs and outputs. Cooper and Tone (1997) used simulation to study a combined DEA–regression model. Friedman and Sinuany-Stern (1997) developed a methodology using canonical correlation analysis to provide a full rank scaling for all the units.

Their methodology closed the gap between the frontier approach of DEA with the average tendencies of statistics.

Compared to the methods mentioned, DEA has several advantages. Multiple inputs and outputs can be used in the DEA model. The weights that will be used to aggregate inputs and outputs are determined using linear programming. No decisions need be made regarding the relative importance of each input and output. With DEA, each operating unit's efficiency is compared to an ''ideal'' operating unit rather than to average performance.

DEA has some limitations as well, however. As with any other method of determining efficiency, all inputs and outputs must be specified and measured. Failure to include a valid input or output or inclusion of an invalid input or output will bias the results. Additionally, DEA can measure ''relative'' efficiency, but not ''absolute'' efficiency. It compares an operating unit to a subset of peers and not to a theoretical maximum performance.

4. APPLICATIONS OF DEA TO THE HEALTH CARE INDUSTRY

DEA has been extensively applied to the health care industry. Wagner and Shimshak (2000) used DEA to evaluate the efficiency of primary care physicians from a managed care organization. DEA was used to measure the relative technical efficiencies of 164 HMOs licensed to practice in the United States in 1995. The data in this study was collected from the American Association of Health Plans (Siddharthan, Ahern, & Rosenman, 2000). Rahman (2007) conducted a comparative efficiency analysis of health clinics in Bangladesh and community health clinics in the San Joaquin Valley in California using DEA. A DEA model that can be implemented by public sector management for assessing the efficiency of a health system within a developing country was developed by Alexander, Busch, and Stringer (2003). A DEA model for determining the operational efficiency of each type of hospital found in Taiwan and for identifying an improved way of allocating resources was developed by Hsu, Pi-Fang, and Hui-Chen (2007). DEA was used by Dexter and O'Neill (2004) to determine by how much hospitals could increase elective inpatient surgical workload for various specialties. Kirigia, Emrouznejad, Sambo, Munguti, and Liambila (2004) measured technical efficiency of public health centers in Kenya using DEA. Various types of DEA applications to the health care industry were described by Chilinqerian and Sherman (2004).

5. RESULTS AND ANALYSIS OF EVALUATING THE RELATIVE EFFICIENCY OF GYNECOLOGICAL DEPARTMENTS

Two sets of DEA were performed using the data obtained from the 2007 U.S. News & World Report study. The objective of the first set was to determine the impact of service resources on the efficiency scores, while the objective of the second set was to determine the impact of nursing resources on the efficiency scores. Service resources include advanced services and patient services. Advanced services include full-field digital mammography, infection isolation rooms, positron emission tomography scanners, and stereotactic radio surgery. Patient services include fertility clinics, genetic testing and counseling, hospices, pain management programs, palliative care, patient-controlled analgesia, rehabilitation care, and translators. The advanced services and patient services data are measure of how many of these services a particular hospital has. This input group measures the hospital's advanced technical resources and patient programs. The other input set is a personnel measure, particularly nursing resources. The nursing index is the relative ratio of nurses to patients, and a higher number is better. A nurse magnet hospital satisfies standards set by the American Nurses Credentialing Center (U.S. News & World Report, 2008). These are the resources or inputs we will take into consideration in our analysis. Identical output was used in both sets (i.e., mortality index and reputation). The mortality index reflects the relative ratio between actual deaths and expected deaths – a smaller ratio is better. For example, a hospital that was expected to have 50 deaths, but only had 35 deaths would have a 0.70 mortality index. Reputation is obtained from peer-based surveys administered by U.S. News &World Report. The mortality index data given in U.S. News & World Report had to be converted in order to be consistent with the notion that the higher the score, the better it was. This was accomplished by subtracting each datum entry given in U.S. News & World Report from 1. The data regarding nurse magnet hospitals was in the form of "yes" or "no." To allow the use of these data in the DEA model, a "yes" was replaced by "1" and a "no" by "0."

The DEA model represented by relations (1)–(5) was utilized. Service resource efficiency scores are summarized in Table 1 and nursing resources efficiency scores are summarized in Table 2. Table 3 summarizes the comparative ranking of U.S. News & World Report Ranking with DEA-based ranking of both service resource efficiency and nursing resource efficiency. The hospital gynecological departments listed in Table 3 are

Table 1. Summary of Service Resources Efficiency Scores for the Top 20 Gynecological Hospital Departments in the United States.

Hospital Gynecological Department	Input Data		Output Data		Service Resource Efficiency Score
	Advanced services	Patient services	Mortality index	Reputation	
UCLA Medical Center	4	4	0.48	7.4	1
University of Utah	3	7	0.32	4.8	1
Stanford Hospital	4	6	0.25	5.4	1
Duke University Medical Center	4	7	0.82	14	1
UCSF Medical Center	3	6	0.34	13.5	1
Parkland Memorial Hospital Dallas	2	7	1.19	11.8	1
John Hopkins Hospital	3.5	8	0.47	26	1
University of Alabama Hospital	2.5	6	0.75	5.8	1
Brigham & Women's Hospital, Boston	4	7	0.34	21.5	0.863
Magee-Womens Hospital of UPMC	3	7	0.36	8.8	0.988
University of Washington MC	3.5	7	0.26	7	0.984
University of Texas Anderson Cancer Center	4	5	0.51	11.5	0.976
Vanderbult University MC	4	7	0.23	7	0.941
Memorial-Sloan Kettering Cancer Center	4	6	0.36	5.1	0.895
Cleveland Clinic	3.5	7	0.35	10.6	0.856
New York-Presbyterian University Hospital of Columbia and Cornell	4	8	0.73	12.2	0.854
Mayo Clinic	4	8	0.57	17.7	0.822
Yale-New Haven Hospital	4	7	0.41	7.9	0.794
Massachusetts General Hospital	4	7	0.79	8.9	0.767
University of Pennsylvania Hospital	4	8	0.38	5.3	0.744

ordered identically to the way in which they were ordered in U.S. News & World Report (i.e., John Hopkins Hospital being placed at the top).

Analysis of the results summarized in Tables 1 and 2 and the solutions to the LP problems formulated by relations (1)–(5) suggest that gynecological departments having an efficiency score of 1 also had zero slacks and were, therefore, Pareto–Koopmans efficient (see Charnes, Cooper, Lewin, Morey, & Rousseau, 1985b; Chang & Kao, 1992).

As can be seen from Table 1, eight gynecological departments had service resource efficiency scores of 1. These departments belonged to the following hospitals: University of California Los Angeles Medical Center, University of Utah, Stanford University, Duke University Medical Center, University

Table 2. Summary of Nursing Resources Efficiency Scores for the Top 20 Gynecological Hospital Departments in the United States.

Hospital Gynecological Department	Input Data		Output Data		Nursing Resource Efficiency Score
	Nursing index	Nurse magnet hospital	Mortality index	Reputation	
John Hopkins Hospital	1.9	Yes	0.47	26	1
Brigham & Women's Hospital, Boston	2.3	No	0.34	21.5	1
Duke University Medical Center	1.6	No	0.18	14	1
New York-Presbyterian University Hospital	1.7	No	0.27	12.2	1
University of Pennsylvania Hospital	1.5	No	0.38	5.3	1
Memorial-Sloan Kettering Cancer Center	1.5	No	0.36	5.1	1
Parkland Memorial Hospital	1.7	No	1.19	11.8	0.993
Magee-Womens Hospital of UPMC	1.8	No	0.36	8.8	0.89
Vanderbult University MC	1.8	Yes	0.23	7	0.893
Stanford Hospital	1.8	Yes	0.25	5.4	0.867
University of Utah Hospital	1.9	No	0.32	4.8	0.836
University of Texas Anderson Cancer Center	1.9	Yes	0.51	11.5	0.827
UCSF Medical Center	2.2	No	0.34	13.5	0.81
Cleveland Clinic	2	Yes	0.35	10.6	0.78
Massachusetts General Hospital	2	Yes	0.79	8.9	0.77
University of Washington MC	2.1	Yes	0.26	7	0.74
University of Alabama Hospital	2.1	Yes	0.75	5.8	0.717
UCLA Medical Center	2.4	Yes	0.48	7.4	0.635
Yale-New Haven Hospital	2.5	No	0.41	7.9	0.63
Mayo Clinic	2.8	Yes	0.58	17.7	0.62

of California San Francisco Medical Center, Parkland Memorial Hospital Dallas, John Hopkins Hospital, and University of Alabama Hospital. Those gynecological departments had the best mortality index and reputation scores relative to their number of advanced services and patient services. The other 12 departments were found to be inefficient as they had resource efficiency scores ranging from 0.988 to 0.744. Those gynecological departments were not as efficient at lowering their mortality index and raising their reputation relative to their advanced services and patient services as the previously mentioned departments. The results summarized in Table 2 indicate that six gynecological departments had perfect nursing

Table 3. Comparative Ranking of the Top 20 Gynecological Hospital Departments in the United States.

Hospital Gynecological Department	U.S. News & World Report Ranking	Services Efficiency Ranking	Nursing Efficiency Ranking
John Hopkins Hospital	1	1	1
Brigham & Women's Hospital	2	9	1
Mayo Clinic	3	17	20
Duke University Medical Center	4	1	1
New York Presbyterian Hospital of Cornell and Columbia	5	16	1
UCSF Medical Center	6	1	13
Cleveland Clinic	7	15	14
University of Texas Anderson Cancer Center	8	12	12
University of Washington Medical Center	9	11	16
Vanderbilt University Medical Center	10	13	9
Yale-New Haven Hospital	11	18	19
Magee-Women's Hospital of UPMC	12	10	8
UCLA Medical Center	13	1	18
Massachusetts General Hospital	14	19	15
Parkland Memorial Hospital	15	1	7
Stanford Hospital	16	1	10
University of Utah Hospital	17	1	11
University of Pennsylvania	18	20	1
Memorial Sloan-Ketering Cancer Center	19	14	1
University of Alabama Hospital	20	1	17

resource efficiency scores of 1. Those departments belonged to the following hospitals: John Hopkins Hospital, Brigham & Women's Hospital, Duke University Medical Center, New York-Presbyterian University Hospital, University of Pennsylvania Hospital, and Memorial-Sloan Kettering Cancer Center. Those gynecological departments had the best mortality index and reputation scores relative to their nursing resources. The other 14 departments were found to be inefficient as they had nursing efficiency scores ranging from 0.993 to 0.622. These gynecological departments were not as efficient at lowering their mortality index and raising their reputation relative to their nursing resources as the previously mentioned departments. Only two gynecological departments were found to have both a perfect service and perfect nursing resource efficiency of 1. These departments belonged to John Hopkins Hospital and Duke University Medical Center.

The DEA efficiency scores should be carefully interpreted as a gynecological department that earned a low service and/or nursing resource efficiency score could nonetheless provide outstanding health care service to patients. The efficiency score only means that a given department could obtain a better mortality index and reputation given its resources. The gynecological department at Mayo Clinic, for example, obtained both low service and low nursing resource efficiency scores despite high scores for both service and nursing resources. A possible reason for Mayo Clinic's low efficiency score is that given its reputation and resources, some very sick patients go there as a last resort. Those patients contribute to lowering Mayo Clinic's mortality index.

The mortality index can be truly indicative of gynecological departments only if the severity of patient sickness is taken into consideration. This would necessitate the establishment of a sickness severity scale. Establishment of such a scale, however, would likely be opposed by both the medical and legal professions.

6. CONCLUSIONS

Widely available rankings of health care units and educational programs published by U.S. News & World Report and other national magazines often influence the decisions of patients/students regarding where to obtain service. Such rankings also affect an institutions ability to attract highly qualified professionals. Rankings based on DEA, however, more completely and accurately represent organizational units like gynecological departments and are, therefore, more informative than most rankings provided in trade magazines,

In this study, DEA was used to analyze the relative efficiency of the top 20 hospital gynecological departments ranked by the 2007 U.S. News & World Report. Two sets of DEA were performed. The objective of the first set was to determine the impact of service resources on the efficiency score, and that of the second set was to determine the impact of nursing resources on the efficiency score. The mortality index and the reputation of each department were used as output in both sets. The higher the efficiency score for a gynecological department, the higher the potential of that department to obtain a more favorable mortality index and reputation given its resources.

The comparative rankings of the top 20 gynecological hospital departments in the United States are provided in Table 3. With this table, it is possible to compare the U.S. New & World Report rankings with the DEA-based rankings while also considering service efficiency and nursing efficiency. The DEA-based ranking accrued by each hospital coupled with the efficiency score can serve to motivate and guide hospital administrators so that they may better utilize expensive resources and thereby improve the performance of hospital gynecology departments. Additionally, DEA-based rankings enable the general public to have better insight into how the nation's top gynecology departments compare. It should be noted, however, that the mortality index should not be considered as seriously as it may not always be a true indicator of patient care. The mortality index can be truly indicative of gynecological departments only if the severity of patient sickness is taken into consideration. This would necessitate the establishment of a sickness severity scale. Establishment of such a scale, however, would likely be opposed by both the medical and legal professions.

REFERENCES

Alexander, C. A., Busch, G., & Stringer, K. (2003). Implementing and interpreting a data envelopment analysis model to assess the efficiency of health systems in developing countries. *IMA Journal of Management Mathematics*, *14*(1), 49–63.

Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, *30*(9), 1078–1092.

Center for Disease Control. (2008). *National Center for Health Statistics* (Vol. 11), Fast Stats.

Chang, K. P., & Guh, Y. Y. (1991). Linear production functions and the data envelopment analysis. *European Journal of Operational Research*, *52*, 215–223.

Chang, K. P., & Kao, P. H. (1992). The relative efficiency of public versus private municipal bus firms: An application of data envelopment analysis. *The Journal of Productivity Analysis*, *3*, 67–84.

Charnes, A., & Cooper, W. W. (1985). A preface to topics in data envelopment analysis. *Annals of Operations Research*, *2*, 59–94.

Charnes, A., Cooper, W. W., Golany, B., Seifoed, L., & Stutz, J. (1985a). Foundations of data envelopment analysis for Pareto-Koopmans efficient empirical production functions. *Journal of Econometrics*, *30*, 91–107.

Charnes, A., Cooper, W. W., Lewin, A. Y., Morey, R. C., & Rousseau, J. (1985b). Sensitivity and stability analysis in DEA. *Annals of Operations Research*, *2*, 139–156.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, *2*, 429–444.

Chilinqerian, J. A., & Sherman, H. D. (2004). Health care applications: From hospitals to physicians, from productive efficiency to quality frontiers. In: W. William, L. M. Seiford & J. Zhu (Eds), *Handbook on data envelopment analysis*. New York: Springer Publishers.

Colbert, A., Levary, R. R., & Shaner, M. C. (2000). Determining the relative efficiency of MBA programs using DEA. *European Journal of Operational Research*, *125*, 656–669.

Cook, W. D., & Kress, M. (1990). A data envelopment model for aggregating preference ranking. *Management Science*, *36*(11), 1302–1310.

Cooper, W. W., Seiford, L. M., & Ton, K. (2006). *Introduction to data envelopment analysis*. New York: Springer.

Cooper, W. W., & Tone, K. (1997). Measures of inefficiency in data envelopment analysis and stochastic frontier estimation. *European Journal of Operational Research*, *99*, 72–88.

Cruise, P. L., & Nyhan, R. C. (2000). First among (un)equals: Assessing hospital performance using data envelopment analysis. *Journal of Health & Human Services Administration*, *22*, 354–373.

Dexter, F., & O'Neill, L. (2004). Data envelopment analysis to determine by how much hospitals can increase elective inpatient surgical workload for each specialty. *Anesthesia & Analgesia*, *99*, 1492–1500.

Friedman, L., & Sinuany-Stern, Z. (1997). Scaling units via the canonical correlation analysis in the DEA context. *European Journal of Operational Research*, *100*, 629–637.

Hsu, H., Pi-Fang, C., & Hui-Chen, H. (2007). Development and application of a modified data envelopment analysis for assessing the efficiency of different kinds of hospitals. *International Journal of Management*, *6*, 21–28.

Kirigia, J. M., Emrouznejad, A., Sambo, L. G., Munguti, N., & Liambila, W. (2004). Using data envelopment analysis to measure the technical efficiency of public health. *Journal of Medical Systems*, *28*(2 April), 155–166.

Rahman, M. (2007). A data envelopment analysis in measuring efficiency of the community health clinics: A comparative study between rural Bangladesh and San Joaquin Valley in California. A paper presented at American Public Health Association Annual Meeting, November, Washington, DC.

Seiford, L. M. (1996). Data envelopment analysis: The evolution of the state of the art 1978–1995. *The Journal of Productivity Analysis*, *7*, 99–137.

Sexton, T. R. (1986). The methodology of data envelopment analysis. In: R. H. Silkman (Ed.), *Measuring efficiency: An assessment of data envelopment analysis*. San Francisco, CA: Jossey-Bass.

Siddharthan, K., Ahern, M., & Rosenman, R. (2000). Data envelopment analysis to determine efficiencies of health maintenance organizations. *Health Care Management Science*, *3*(1), 23–29.

Thanassoulis, E., & Dyson, R. G. (1992). Estimating preferred target input-output levels using data envelopment analysis. *European Journal of Operational Research*, *56*, 80–97.

U.S. News & World Report. (2008). *Best hospitals 2007 specialty: Gynecology*, 30 January.

Wagner, J. M., & Shimshak, D. G. (2000). Physician profiling using data envelopment analysis: A case study. *International Journal of Health Care Technology and Management*, *2*(1–4), 358–374.