Journal of Productivity Analysis, 11, 169–182 (1998)

Analysis of Intertemporal Efficiency Trends Using Rank Statistics With an Application Evaluating the Macro Economic Performance of OECD Nations

P. L. BROCKETT
Gus S. Wortham Chair in Risk Management and Insurance, Department of Management Science and Information Systems, The University of Texas at Austin, Austin, TX 78712

B. GOLANY
Faculty of Industrial Engineering and Management, The Technion - Israel Institute of Technology, Haifa, 32000, Israel

SHAN LI
Center for Management of Operations and Logistics, The University of Texas at Austin, Austin, TX 78712

Abstract

This paper presents two applications of rank statistics to evaluate efficiency performance trends using productive efficiency measures derived through various Data Envelopment Analysis (DEA) models. The paper starts with a discussion of the difficulties in obtaining consistent ranks from DEA efficiency ratings. Next, a procedure is proposed to identify intertemporal performance trends using any one of several possible efficiency measures. Another procedure is then developed to test the stability over time of the rank positions of the analyzed units. For each statistical procedure, a small numerical example involving DEA efficiency measures is provided to illustrate the proposed technique. Finally, the new procedures are applied to data reflecting the macro-economic performance of 17 OECD nations in 1979–1988. The outcomes of the application are discussed and contrasted with previous research in this area.

Keywords: Ordinal rank statistics, Intertemporal analysis, Data envelopment analysis, Macro-economic productivity indices

1. Introduction

Since the introduction of the first Data Envelopment Analysis (DEA) model by Charnes, Cooper and Rhodes (CCR) (1978), the literature on the subject has proliferated rapidly and many application areas have been recorded (see Seiford (1995)). These applications have been mainly in the form of pilot projects or one-time studies related to theoretical extensions of the methodology. Only in a few cases (e.g., public schools and Army recruiting units) were attempts made to establish an *on-going* evaluation system based on DEA. For such systems to be successful, the analytical engine of DEA has to be incorporated within an information system containing a full array of data collection, storage, transfer and reporting capabilities (see the application guidelines in Golany and Roll (1989)). Such monitoring systems will regularly record the resources used and the results achieved by a number of Decision Making Units (DMUs) over time, and will then report them to users in various echelons of management.

In contrast to current *re-engineering* trends, such DEA systems are expected to lead towards *continuous* improvement. Practically, this means that at the end of each time period, the data pertaining to it is incorporated with existing data (possibly replacing the data corresponding to the most dated period—as in a moving window) and evaluated by the DEA methodology. Analysis across *time*, rather than within individual periods, is much more appropriate for decision making. Managers are reluctant to risk making a decision on the basis of only the most current period since, whatever has been observed, the cause might be an unusual set of circumstances that will not occur again. However, when an outcome (e.g., a particular trend) is shown to hold over several periods, it is more likely to be relied upon in such on-going evaluation environments.

In this paper, we discuss two categories of intertemporal studies that can be addressed within such DEA systems. First, we seek to determine efficiency trends in the performance of the DMUs over time. Such trends may develop slowly, sometimes unevenly across different units, and may go unnoticed by the managers of the individual DMUs. Intertemporal analysis offering early detection of 'global' trends may be crucial to the well being of the organization where the DEA system is implemented. We are able to state whether or not there are statistically significant trends in the efficiency ratings over time.

The second task is to analyze the stability of the individual DMUs relative rankings over time. The question here is whether certain DMUs always dominate others regardless of any overall performance trend. This analysis can be conducted either in a contemporaneous or intertemporal manner. In the contemporaneous approach, one would evaluate each time period separately, record the resultant rankings for each DMU and later compare it with other rankings for other DMUs that were generated in a similar manner. In the intertemporal approach, DMUs corresponding to a number of time periods are run simultaneously and their efficiency scores are transformed into rankings across the periods. Different statistical techniques apply to these two different approaches resulting in possibly different interpretations for the results.

This paper extends an earlier work (cf., Banker (1984), Brockett and Golany (1996), and Pastor et al. (1996)) in which non-parametric rank statistical approaches were used in the DEA context.[1] In particular, Brockett and Golany (1996) developed a statistical significance test for 'program evaluation' problems in DEA using rank statistical methods. In their paper they gave several reasons that motivate the use of rank statistics instead of parametric statistical procedures based on the efficiency ratings themselves. First, the non-parametric statistical analysis is more consistent with the non-parametric approach to production function determination that originally motivated the invention of DEA efficiency analysis. Second, the statistical distribution of efficiency scores are generally not known, however, by translating the efficiency outcomes to their corresponding ranks we have access to non-parametric and robust statistical test statistics which do have known statistical distributions and asymptotics. Thus, describing efficiency values using rank statistics bypasses the difficulty inherent in the parametric models of efficiency scores. Third, different DEA model formulations (and different efficiency frontier techniques) applied to the same data set can produce efficiency ratings in different ranges and spreads, making the parametric approach more susceptible to the choice of model assumed for analysis. By replacing the efficiency ratings with their respective rankings comparisons of efficiency distributions across different DEA models are facilitated. The reader is referred to Ruefli (1990, Chapter 2) for a more detailed discussion of the advantages and problems that may arise when ordinal rather than cardinal data is used in the analysis of time series.

In the application section of this paper, we rank order the units based upon their respective distances from the efficiency frontier. To do so, we use the largest value r by which the inputs can be adjusted without changing the efficiency classification of the unit (cf., Charnes et al. (1992)). This distance measure, also known as the "radius of stability" provides just one method of ranking, and the analysis presented in the paper could equally well be applied to the ranks derived from the original DEA efficiencies, or from any alternative scoring method.

To illustrate the techniques proposed in this paper, we analyze the relative efficiency performance over time for a sample of 17 Organization for Economic Cooperation and Development (OECD) countries over the period 1979–1988. Employing the proposed rank statistic approach, we are able to determine if there are significant trends in the efficiency patterns of each country over time. Furthermore, we can ascertain whether it can be said, with statistical confidence, that the 17 OECD countries have maintained their relative efficiency rank positions in the group over time (i.e., determining whether some countries are consistently outperforming the others).

A brief summary of the paper follows. In §2 we construct the framework for a variety of intertemporal analyses and provide appropriate non-parametric rank-based statistical tests to check for the existence of trends in efficiency performance over time. A small numerical example is used to illustrate the technique. In §3, we show how to use the same mathematical structure developed earlier to investigate whether the individual DMUs maintain their relative efficiency positions vis-a-vis their peer group, and we explain the managerial implications of this proposed statistical test. To demonstrate the test procedure, we revisit the example from §2. In §4 we illustrate the proposed procedures with an analysis of the OECD nations macro-economic performance. Final conclusions and directions for further research are offered in §5.

2. Detecting Trends Over Time

When consecutive observations corresponding to the performance of a group of DMUs over time are evaluated simultaneously, it is important to detect if there are significant trends in the efficiency ratings or in other key variables. Such intertemporal analysis within the DEA context was initiated in a 'window analysis' study for the U.S. Army Recruiting Command (Ali et al. (1981)). However, this work, as well as subsequent 'window' studies (e.g., Charnes et al., 1985), did not lead to further statistical developments in testing for specific patterns in the results.

We develop this statistical analysis by building on a model structure and approach introduced by Brockett and Kemperman (BK) (1980) for non-parametric detection of trends

Table 1. Efficiency ratings matrix.

Period(t)\DMUs(j)	1	...	n
1	h_{11}		h_{1n}
...		h_{tj}	
K	h_{k1}		h_{kn}

Table 2. Column rankings of the efficiency matrix.

Period(t)\DMUs(j)	1	...	n
1	C_{11}		C_{1n}
...		C_{tj}	
K	C_{k1}		C_{kn}

in "fuzzy" or "dirty" data. The BK paper was designed to ascertain whether or not there were trends in the adverse reactions to drugs in data reported to the World Health Organization. In order to access the known strengths and robustness of non-parametric rank statistics, they introduced a "rank matrix transformation" which we shall use in this paper to transform the original non-metric-quality data into rank ordered data. They then showed how (essentially) a correlation between the time index and the rank value could be used to detect trends. To adapt this methodology to the setting of the current paper, we start with the matrix of observed efficiencies as presented in Table 1.

The efficiency ratings in Table 1 (h_{tj}, $t = 1, \ldots, k$; $j = 1, \ldots, n$) were generated by performing a single DEA run in which $n \cdot k$ DMUs were evaluated, one for each of the n DMUs in each of k time periods. The use of a single efficiency frontier for the evaluation of all $n \cdot k$ observations implicitly assumes (as per the null hypothesis) that no technological changes affecting productive efficiency have occurred over the k time periods.

We note that the efficiency ratings across time may result from comparisons to different facets consisting of DMUs taken from different periods. The same is true for different DMUs in the same period. Accordingly, as explained in §1, we do not focus our statistical analysis on the actual DEA derived efficiency ratings (since their exact metric qualities are unknown and their precise metric value may depend on the DEA model used). Instead, for the purpose of identifying trends, we prefer to observe the relative *ranking* of the ratings for each DMU across time.

Following the methodology outlined in BK, we replace the actual efficiency ratings in each *column* of the efficiency matrix with the corresponding rank statistic obtained by ordering the scores within the column in an ascending order. In this manner, we obtain the rank-value matrix in Table 2. As in the BK model, when efficiency ties are present, we replace the relevant rankings with their midrank.[2]

The null hypothesis is that the vector of k rankings for each DMU_j is not dependent on time; i.e., that the observed rankings $C_{1j}, C_{2j}, \ldots, C_{kj}$ are exchangeable so that any rearrangement of the elements of such a vector is equally likely to occur. A two-sided alternative hypothesis is posed incorporating both increasing and decreasing trends over time. Of course, one-sided alternative and tests are available if one has a priori reason to suspect a trend in a particular direction. In particular, BK (p. 109) also give an alternative test statistic for detecting an increasing trend over time.

An important assumption that was stated by BK (ibid., p. 108) was that the original observed vectors $\mathbf{h}_{\cdot\mathbf{j}}$ (here h_{tj} considered as a vector over t for DMU$_j$) be independent of $\mathbf{h}_{\cdot\mathbf{k}}$ for j. The internal components of the vector $\mathbf{h}_{\cdot\mathbf{j}}$ may, however be dependent for fixed j. This issue, in the context of the DEA efficiency evaluation, was investigated by Banker (1993) who postulated that the efficiency deviations can be considered as statistically independent of one another (ibid., p. 1268, Postulate 3A).

Under the null hypothesis, the rank vector for any DMUj $(C_{1j}, C_{2j}, \ldots, C_{Tj})$ is uniformly distributed over the set of all $k!$ possible arrangements[3] of $(1, 2 \ldots, k)$; i.e., each possible arrangement has the same probability to occur (namely $1/k!$). Following the analysis of BK we compute the test statistic

$$S = \sum_{j=1}^{n} \sum_{t=1}^{k} t \cdot C_{tj} \tag{1}$$

When ties are allowed to occur and the midranks are used to describe them, BK have shown that for sufficiently large values of $n \cdot k$ (e.g., $n \cdot k \geq 11$) the distribution of S is approximately normal with the mean and variance parameters as follows:

$$\begin{aligned} \mu &= E(S) = \frac{n \cdot k}{4} \cdot (k+1)^2 \\ \sigma^2 &= Var(S) = \frac{n}{144} \cdot k^2 \cdot (k^2 - 1) \cdot (k+1) - \frac{k \cdot (k+1)}{144} \cdot \sum_{j=1}^{n} \sum_{r} (d_{jr}^3 - d_{jr}) \end{aligned} \tag{2}$$

In this analysis, the term d_{jr} denotes the number of values $h_{1j}, \ldots, h_{kj}$ that are tied in the r$^{\text{th}}$ group of ties (when there are relatively few ties the second term in the variance computation can be ignored). The number of DMUs in any DEA application is such that the preceding condition is guaranteed to hold for any $k \geq 2$.

The next step is to transform the test statistic S to an approximately standard normal distribution by using $Z = (S - \mu)/\sigma$. The null hypothesis that there is no trend in the observed efficiencies can be rejected against a two sided alternative hypothesis of trend at a level of significance α whenever $Z \leq -Z_{\alpha/2}$ or $Z \geq Z_{\alpha/2}$ where $Z_{\alpha/2}$ denotes the upper $\alpha/2$ percentile of the normal distribution. Similarly, the null hypothesis of no trend in the observed efficiencies can be rejected against a one sided alternative hypothesis of an increasing (decreasing) trend at a level of significance α whenever $Z \geq Z_\alpha$ (respectively $Z \leq -Z_\alpha$).

Numerical Example. Consider the efficiencies of 5 DMUs over 4 periods and their respective rankings[4] as presented in Tables 3 and 4. Intuitively, we observe that most DMUs exhibit an upward trend in their ratings. However, DMUs 3 and 4 in Table 4 do not always conform to this conjectured trend. Now, with these mixed signals, we wish to investigate if we have enough evidence to reject the null hypothesis that no general trend exists against a two sided alternative hypothesis. Using equations (2) and (3), the parameters

Table 3. Efficiency matrix (example).

$t \backslash j$	1	2	3	4	5
1	.87	.92	.96	.95	.91
2	.93	1.0	.95	.97	.92
3	.96	1.0	.94	.95	.94
4	1.0	1.0	.98	.99	.96

Table 4. Column ranking matrix (example).

$t \backslash j$	1	2	3	4	5
1	1	1	3	1.5	1
2	2	3	2	3	2
3	3	3	1	1.5	3
4	4	3	4	4	4

of the approximated normal distribution for this case are $\mu = 125$, $\sigma^2 = 37.5$, the value of the test statistic is $S = 142$ and the corresponding value from a standard normal distribution is $Z = 2.776$. Hence, the null hypothesis can be rejected for this example with a confidence level larger than 99.7% and we can conclude that some trend over time is present.

3. Detecting Stable Efficiency Rankings

The same information gathered in Table 1 can be used for another important aspect of the intertemporal analysis. Ranking the set of $n \cdot k$ efficiency ratings and observing the sum of ranks associated with each DMU, one can learn about the relative position of the DMUs vis-á-vis each other across the whole period (not necessarily in a particular time). Whereas in the previous section we analyzed whether the performance of the entire group deteriorates, improves or stays the same over time, here we investigate whether it can be said, with statistical confidence, that the n DMUs all maintained their same relative position in the group over time. Previously this hypothesis was difficult to test statistically since typically most DMUs experience some ups and downs in their relative rankings.

To provide an analytical answer to the stability question we apply here the Kruskal-Wallis nonparametric ANOVA test (cf., Brockett and Levine (1984)). For this test there are n ‘populations’ simultaneously under investigation and the null hypothesis is that all n populations (DMUs in our application) have the same distribution of ratings. To apply this statistical methodology we first rank order the set of $n \cdot k$ scores in an ascending order (ties are again broken by the midrank) and let R_j denote the sum of the ranks corresponding to DMU_j. Then, we compute the Kruskal-Wallis test statistic as:

$$H = \frac{12}{n \cdot k \cdot (n \cdot k + 1)} \cdot \left(\frac{R_1^2}{k} + \frac{R_2^2}{k} + \cdots + \frac{R_n^2}{k} \right) - 3 \cdot (n \cdot k + 1) \tag{3}$$

Table 5. Sum of ranks matrix (example).

J	1	2	3	4	5
R_j	36.5	59	42.5	48	24

It is known that this test statistic is distributed according to a χ^2 distribution with $n - 1$ degree of freedom (cf., Brockett and Levine (1984)). Rejection of the null hypothesis leads to the conclusion that, in general, the different DMUs maintain their relative efficiency positions over time; i.e., at least one DMU is consistently better or worse than the others. When the null hypothesis is rejected, a multiple comparison test can be performed to gain further insights into the stable structure of the rankings.

The rejection of the null hypothesis indicates there are stable differences which occur among the DMUs over time. An investigation of why this is so may have significant managerial impact for several reasons. One possible explanation for the rejection which should be considered is that the set of inputs and outputs which were selected for the DEA model lacked certain important factors that participate in the process under investigation (i.e., the model was incorrectly constructed). For example, there might be some environmental factor (e.g., a certain demographic characteristic) that helps particular DMUs generate consistently more outputs while using the same amounts of discretionary inputs as other DMUs. When this is observed over a sufficiently long time in which some managers were replaced, management practices were altered, etc., we may suspect that the phenomenon is inherent to the 'production process' and not due to the quality of management. On the other hand, accepting the null hypothesis amounts to (at least a partial) *validation* of the input-output formulation in DEA.

However, when we have enough confidence that the model has been correctly specified and no hidden factors have been left out, rejection of the null hypothesis can be used to justify the selection of consistent 'winners' as sites for future marketing tests, pointing out consistent 'losers' as candidates for closure, evaluating the success of training programs which were aimed at bringing all the DMUs up to par, etc.

Numerical Example. Returning now to the numerical example given in §2, we rank order the set of 20 efficiency scores and compute the R_j values. These values are presented in Table 5.

The value of the Kruskal-Wallis test statistic that corresponds to these values is $H = 4.853$. This value of H when compared to $\chi^2_4(0.05) = 9.49$ does not allow the rejection of the null hypothesis. In fact, the hypothesis of equivalent distribution of efficiency ranking for all five DMUs cannot be rejected for any reasonable level of significance.

4. An Intertemporal Analysis of the Macroeconomic Performance of the OECD Countries

We now apply the previously developed techniques to the analysis of productivity growth for a sample of Organization for Economic Cooperation and Development (OECD) countries.

The motivation to select this application stems from recent DEA studies of macro-economic data in which the authors took part (see the study of the G-7 countries in Golany and Thore (1997a), and the studies of the OECD countries in Golany and Thore (1997b) and in Lovell et al. (1995)). The data in this particular application is the same as that used by Färe et al. (1994) who had the same purpose as the one we have in mind, namely—identifying productivity trends. We used a sample of data on $n = 17$ OECD countries[5] over the period 1979–1988 obtained from the Penn World Tables (Mark 5) which was built from the International Comparison Program of the United Nations and national account data. The procedures used to create the data set are discussed in some detail in Summers and Heston (1991). The resulting adjustments imply that "real international quantity comparison can be made both between countries and over time" (Summers and Heston (1991, p. 1)). The international prices are average world prices of final goods, rather than prices of a specific benchmark country.

Intertemporal analysis of the type performed here (and in Färe et al. (1994)) sheds light on issues related to the hypothesized productivity slowdown observed in the United States and other industrialized countries during the 1970s and 1980s, a topic which has received great attention. The implications of this phenomenon for the competitive position of the United States, especially relative to Japan, have become a matter of public debate. Many studies have been devoted to determining whether this relative slowdown is part of a natural longer-term pattern of productivity convergence among countries. Under this hypothesis, such productivity trends are viewed as a natural process of convergence, as countries with low initial levels of productivity exploit the public-goods aspects of technology advances.

The convergence view has been discussed by many researchers, including Abramovitz (1986, 1990), Baumol (1986), and Baumol et al. (1989). Using data collected by Maddison (1982, 1989), these authors provide evidence that incomes have been converging over a fairly long period. Dowrick and Nguyen (1989) have added further evidence for convergence based on the postwar period for a sample of OECD countries. They argue that one needs to distinguish between catch-up or convergence of income (measured by income per capita or income per work hour) and total factor productivity (TFP) catch-up. Following Baumol (1986) and Abramovitz (1986, 1990), Dowrick and Nguyen concluded that TFP catch-up is inversely related to a country's initial level of relative labor productivity. Dowrick (1989) extended the Dowrick and Nguyen results by allowing for sectoral changes. He found evidence that "GDP growth since 1950 has been systematically higher in those OECD countries which have been able to reallocate a greater proportion of their labor force out of agriculture" (p. 335).

Färe et al. (1994) used an alternative measure (the Malmquist index) of total factor productivity growth (TFP) to provide evidence concerning a pattern of total factor productivity growth (including TFP catch up). The technique they used allow them to decompose productivity growth into two components: changes in technical efficiency over time and shifts in technology over time. These components lend themselves in a natural way to the identification of catching up and the identification of innovation, respectively. However, since Färe et al. used the geometric mean of two Malmquist productivity indexes as their measure of productivity growth, they could only do the pairwise comparison analysis over each pair

Table 6. Matrix of countries rankings from the productivity indices matrix.

Country/year	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988
Australia	10	8	6	2	1	3	4	5	7	9
Austria	10	9	8	7	6	4.5	1.5	3	1.5	5
Belgium	9	10	8	6.5	5	3.5	1	2	4.5	6.5
Canada	5	2	3	1	4	6	7	8	9	10
Denmark	4	2	1	3	5.5	8	10	9	7	6.5
Finland	1	2	3	4	5	6	7	8	9	10
France	7.5	7.5	4.5	6	3	2	1	4.5	9	10
Germany	7	9	7	4	2	3	1	5	7	10
Greece	10	9	8	5.5	3	1	4	6.5	2	7
Ireland	10	9	8	7	6	5	4	1	3	2
Italy	1	7.5	6	3	2	4.5	4.5	7.5	9	10
Japan	1	2	3	4	5	6	7	8	9	10
Norway	5	3.5	2	1	3.5	6	7	8	9	10
Spain	10	9	8	7	6	4	1	2	3	5
Sweden	5	3.5	1.5	1.5	3.5	9	10	7	8	6
U.K.	10	3	1	2	5	4	6	7	9	8
U.S.A.	9	4	3	1	2	6.5	5	6.5	8	10

of years and consequently a large number (918) of linear programming problems had to be solved to get their results.

This section shows that by using our rank statistic approach, we can determine if there are significant trends in the productivity growth for each country over time and whether it can be said, with statistical confidence, that the 17 OECD countries have maintained their relative positions in the group over time (i.e., to determine whether some countries are consistently outperforming the others with regard to productivity efficiency).

For the purpose of comparison of our results with Färe et al.'s (1994), we have performed our DEA analysis using gross domestic product (GDP) as our measure of aggregate output, and capital stock and employment as our aggregate input proxies.[6] GDP and capital stock are measured in constant 1985 international prices. Employment is found using the real GDP per worker, and capital is obtained from the capital stock per worker. (Capital stock does not include residential construction but does include gross domestic investment in producers' durables, as well as nonresidential construction. These are the cumulated depreciated sums of past investment.)

The DEA results were obtained by running 170 linear programming problems for the set of input-output variables, from which we get[7] the productivity index for each country from 1979 to 1988. These are then transformed, as described in §2, to a corresponding matrix of relative rank statistics obtained by ordering the scores for each country in an ascending order. The resulting rank value matrix is given in Table 6. As explained earlier, when ties were present, we replaced the relevant rank by the midrank.

We first observe that while a few countries seem to exhibit an upward trend in their productivity ratings from 1979 to 1988, most of the countries do not seem to consistently conform to such a trend. In order to investigate if there is sufficient evidence to *statistically* reject the null hypothesis that no trend exists, we use equations (2) and (3) for S, μ and

σ^2 to calculate the parameters of the approximating normal distribution. From Table 6 we calculate the value of the test statistic to be $Z = 1.78$ and the corresponding p-value from a standard normal distribution to be 0.076. In this case, the null hypothesis can not be rejected at the 0.05 level of significance and we conclude that no statistically significant trend in productivity (increasing or decreasing) over time exists for these countries as a whole. That is to say, statistically, on the average, productivity did not significantly increase or decrease over the 1979–1988 period for the countries in our sample.

Note that in calculating the productivity index, we applied DEA to the entire set of $17 \times 10 = 170$ DMUs. This permits us to carry out efficiency comparison over time for the *same* country as well as efficiency comparisons over *different* countries for the same time period. By doing this, we can detect both the productivity growth trends over time for each country and the relative efficiency position of each country vis-a-vis each other across the period. In Färe et al.'s discussion, however, instead of doing the intertemporal comparisons, they construct a world frontier based on the data for each specific year and then each country is compared to that frontier. How much closer a country gets to the world frontier is what they call "catching up"; while how much the world frontier shifts at each country's observed input mix is what they call "technical change" or "innovation". The product of these two components yields a frontier version of productivity change (i.e., their productivity index). It should be noted, however, that the Färe et al. (1994), article tacitly assumes an *interval metric scaling* to the efficiency scores which, as discussed in §1, is probably more ordinal non-interval in characteristic.

The next task is to analyze the stability of the relative productivity growth over time for the individual countries. Ranking the entire set of $n \cdot k = 17 \times 10 = 170$ productivity indices and observing the sum of ranks associated with each country, one can learn about the relative position of the countries vis-a-vis each other across the $k = 10$ year period. Here we want to investigate whether it can be said, with statistical confidence, that the $n = 17$ countries tended to maintain their relative position in the group over time. Typically, it is quite difficult to come up with intuitive assessment concerning the stability of the relative rank positions of the different countries since most countries experience some ups and downs in their relative ranking. To provide an analytical answer to the stability question we apply the Kruskal-Wallis nonparametric ANOVA test to the rank matrix as described in §3. There are $n = 17$ 'populations' or countries in this case and the null hypothesis is that all n populations (countries) have the same distribution of ordinal ratings. We rank the set of $n \cdot k$ scores in an ascending order (ties are again replaced by their midrank) and let R_j denotes the sum of the ranks corresponding to country j. These sums of individual ranks for each country are given in Table 7.

To perform the country comparisons, we compute the Kruskal-Wallis statistic using equation (4):

$$H = \frac{12}{170 \cdot 171} \cdot \left(\frac{R_1^2}{10} + \frac{R_2^2}{10} + \cdots + \frac{R_{17}^2}{10} \right) - 3 \cdot 171$$

This test statistic is distributed according to a χ^2 distribution with $n - 1 = 16$ degree of freedom. Rejection of the null hypothesis leads to the conclusion that, in general, coun-

Table 7. Sum of ranks matrix for country productivity over time.

Country	Australia	Austria	Belgium	Canada	Denmark	Finland
R_j	891	524	1208	787	66	185
Country	France	Germany	Greece	Ireland	Italy	Japan
R_j	827	410	1587	762	1462	214
Country	Norway	Spain	Sweden	U.K.	U.S.A.	
R_j	527	1459	933	1197	1496	

tries maintain their relative productivity growth position over time (i.e., some consistently outperform others over time).

The value of the Kruskal-Wallis test statistic that corresponds to the overall ranks in Table 7 is $H = 157.24$ which, when compared to $\chi^2_{16}(.005) = 34.27$ does allow the rejection of the null hypothesis of identical distribution of efficiency ranking for all 17 countries at a level of significance 0.005. This rejection of the null hypothesis leads to the conclusion that some of the 17 countries exhibit consistently better economic performance than others, as measured by the input-output factors that were included in our model. A non-statistical (graphical) aid can give further insight into the causes for the rejection of the null hypothesis. Figure 1 presents a visual summary of the rank statistical results for each country over time. Moreover, using Table 7, the 17 countries can be ordered from highest to lowest by their overall sum-of-ranks as follows: Greece, U.S., Italy, Spain, Belgium, U.K., Sweden, Australia, France, Canada, Ireland, Norway, Austria, Germany, Japan, Finland, and Denmark, which means Greece exhibited the overall largest average rank (i.e., worst performance) and Denmark the overall lowest average rank (i.e., best performance) for the period 1979–1988.

Turning to a further examination of the country-by-country results, as seen in Figure 1, we observe a steady increase in the productivity growth of Japan and Finland over the period 1979–1988, while the productivity growth of Ireland almost strictly decreased. For the U.S., the productivity growth went up and down over these years. Comparing the relative efficiency positions of U.S. and Japan, as shown in Figure 1, the position of U.S. is consistently higher than the position of Japan over the entire period. By the sum of the overall ranks in Table 7, U.S. is ranked in second place and Japan is ranked in the third to last place among the 17 countries indicating that the U.S. did consistently better in its productivity efficiency than did Japan from 1979 to 1988 even though the productivity growth of U.S. exhibited a 'slowdown' relative to Japan. This conclusion is consistent with Färe et al.'s result for Japan, i.e., Japan is moving towards the frontier but is not on the efficiency frontier.

5. Conclusions

The statistical analysis based on the ranking techniques offered in this paper expand the potential benefit of DEA applications while retaining the robust nonparametric nature of

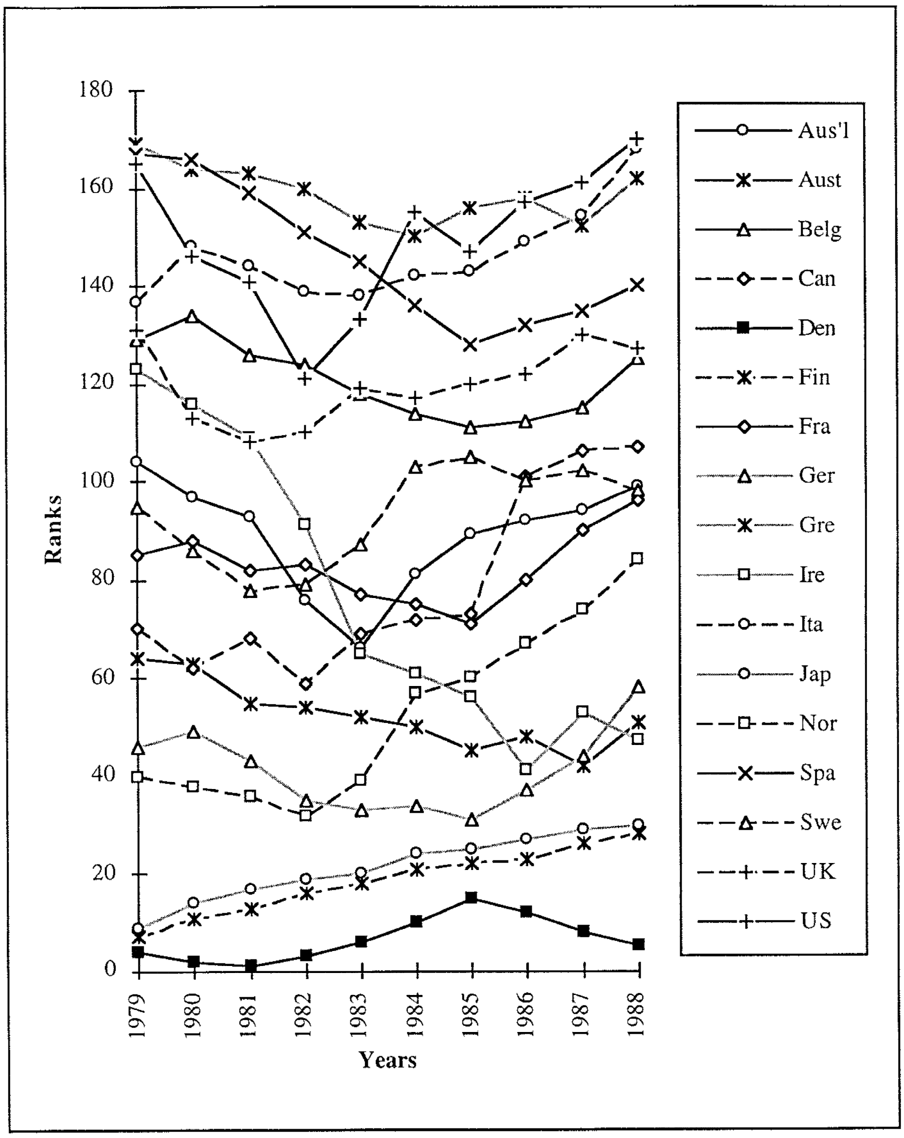

Figure 1. Relative rank positions of countries over time.

the original DEA development. First, proper detection of performance trends over time is an essential element in management monitoring of business units. Second, the study of stability in rank positions may help isolate a smaller group of consistently top performers for the purpose of benchmarking.

As we said all along in this paper, the statistical tests and procedures proposed here are not limited only to DEA. Other areas of ranking theory could be explored as well. For example, the ratio computed between each pair of input-output factors provides a certain ranking of the DMUs. For each DMU that is considered, there are a total of $(m+s)(m+s-1)/2$ such rankings (where $m+s$ is the total number of inputs and outputs). In this context, each DMU is a "judge" providing a pairwise matrix of evaluation of the inputs and the outputs. Combining these judgments into a single aggregate ranking without using a-priori weights is a task similar to the one that DEA performs.[8] The ranking theory, see e.g., Kendal (1962), is replete with such rank aggregation techniques. Further relationships can be established by using other nonparametric rank statistical procedures applied to DEA analyses. These topics shall be investigated subsequently.

Notes

1. The reader is referred to Ruefli (1990, Chapter 2) for a discussion of the advantages and disadvantages that may arise when ordinal rather than cardinal data is used in the more general context of time series analysis.
2. For more details on the commonly used technique of using midranks in cases of tied ranks, see Lehmann (1975). Chapter 1, section 4 of this book gives a very extensive discussion of the use of midranks for the analysis of ranked data and recommends their use in case of ties.
3. This is a conservative estimate since it neglects the possible arrangements that may arise when ties are present.
4. The rankings were obtained by applying DEA to the entire set of $5 \times 4 = 20$ DMUs over time. This allows for efficiency comparison over time for the same DMU, as well as efficiency comparisons over DMUs for the same time period.
5. The 17 OECD countries for which necessary data were available over the period 1979–1988 were Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Japan, Norway, Spain, Sweden, the United Kingdom, and the United States.
6. For more details on these calculations see Li (1995).
7. The distance from the efficiency frontier (sensitivity to changes in inputs needed to change the efficiency classification) was used rather than the raw efficiency scores for creating the productivity index for the reasons discussed in §2. Since only the ordinal characteristics of the actual productivity numbers are used in the statistical analysis, the final analysis is robust to the different version of DEA used or the efficiency score versus distance from the efficiency frontier.
8. See Cook and Kress (1990) for another DEA related ranking technique.

References

Abramovitz, M. (1986). "Catching Up, Forging Ahead and Falling Behind." *Journal of Economic History* 46(2), 385–406.

Abramovitz, M. (1990). "Catching Up Factor in Postwar Economic Growth." *Economic Inquiry* 28(1), 1–18.

Ali, I. A., A. Charnes, W. W. Cooper, D. Divine, and J. Stutz. (1982). "An Application of Data Envelopment Analysis to Management of U.S. Army Recruitment Districts." CCS Research Report no. 436, The University of Texas at Austin.

Banker, R. D. (1993). “Maximum Likelihood, Consistency and Data Envelopment Analysis: A Statistical Foundation.” *Management Science* 39(10), 1265–1274.
Banker, R. D. (1984). “Estimating Most Productive Scale Size using Data Envelopment Method.” *European Journal of Operational Research* 17, 35–44.
Banker, R., A. Charnes, and W. W. Cooper. (1984). “Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis.” *Management Science* 30(9), 1078–1092.
Baumol, W. J. (1986). “Productivity Growth, Convergence and Welfare: What the Long-Run Data Show.” *American Economic Review* 76(5), 1072–1085.
Baumol, W. J., S. Blackman, and W. Wolff. (1989). *Productivity and American Leadership: The Long View*. Cambridge, MA: MIT Press.
Brockett, P. L., and B. Golany. (1996). “Using Rank Statistics for Determining Programmatic Efficiency Differences in Data Envelopment Analysis.” *Management Science* 42(3) (March).
Brockett, P. L., and J. H. B. Kemperman. (1980). “Statistical Recognition of Trends in Health Monitoring Systems.” *Methods of Information in Medicine* 19(2), 106–112.
Brockett, P. L., and A. Levine. (1984). *Statistics, Probability and Their Applications*. W. B. Saunders Publishing Co.
Charnes, A., W. W. Cooper, and E. Rhodes. (1981). “Evaluating Program and Managerial Efficiency: An Application of Data Envelopment Analysis to Program Follow Through.” *Management Science* 27(6), 668–697.
Charnes, A., T. Clark, W. W. Cooper, and B. Golany. (1985). “A Development Study of Data Envelopment Analysis in Measuring the Efficiency of Maintenance Units in the U.S. Air Forces.” *Annals of Operations Research* 2, 95–112.
Charnes, A., S. Haag, P. Jaska, and J. Semple. (1992). “Sensitivity of Efficiency Classifications in the Additive Model of DEA.” *International Journal of System Sciences* 23, 789–798.
Cook, W., and M. Kress. (1990). “Data Envelopment Model for Aggregating Preference Rankings.” *Management Science* 36(11), 1302–1310.
Dowrick, S. (1989). “Sectoral Change, Catching Up and Slowing Down: OECD Postwar Economic Growth Revisited.” *Economic Letters* 31(4), 331–335.
Dowrick, S., and D. Nguyen. (1989). “OECD Comparative Economic Growth 1950–1985: Catch Up and Convergence.” *American Economic Review* 79(5), 1010–1030.
Färe, R., S. Grosskopf, M. Norris, and Z. Zhang. (1994) “Productivity Growth, Technical Progress and Efficiency Change in Industrialized Countries.” *American Economic Review* 84(1), 66–83.
Golany, B., and Y. Roll. (1989). “An Application Procedure for DEA.” *OMEGA* 17(3), 237–250.
Golany, B., and S. Thore. (1997a). “On the Competitiveness of Nations.” In *IMPACT - IC2 Contribution to the Advancement of Technology*. Greenwood Publishing Group.
Golany, B., and S. Thore. (1997b). “Restricted Best Practice Selection in DEA: An Overview with a Case Study Evaluating the Socio-Economic Performance of Nations.” *Annals of Operations Research* 73, 117–140.
Kendal, M. (1962). *Rank Correlation Methods* New York: Hafner.
Lehmann, E. (1975). *NonParametrics: Statistical Mehods Based on Ranks*. San Francisco: Holden Day Inc.
Li, S. (1995). *A New Approach to Sensitivity Analysis of the DEA Models and Their Applications to Ranking and Productivity Growth*. Unpublished Ph.D. Thesis, The University of Texas at Austin.
Lovell, C. A. K., J. T. Pastor, and J. A. Turner. (1995). “Measuring Macroeconomic Performance in the OECD: A Comparison of European and Non-European Countries.” *European Journal of Operational Research* 87, 507–518.
Maddison, A. (1982). *Phases of Capitalist Development*. New York: Oxford University Press.
Maddison, A. (1989). *The World Economy in the 20th Century*. Paris: Organization and Development.
Pastor, J. P., J. L. Ruiz, and I. Sirvent. (1996). “A Statistical Test for Nested Radial DEA Models.” Paper presented at the Second Georgia Productivity Workshop, University of Georgia, Athens, Ga.
Roll, Y., and B. Golany. (1993). “Alternate Methods for Treating Factor Weights in DEA.” *OMEGA* 21(1), 99–109.
Ruefli, T. W. (1990). *Ordinal Time Series Analysis*. Quorum Books.
Seiford, L. M. (1995). “A Bibliography of Data Envelopment Analysis (1978–1995).” In A. Charnes, W. W. Cooper, A. Y. Lewin, and L. M. Seiford (eds.), *Data Envelopment Analysis: Theory, Methodology and Applications*. Kluwer Academic Publishers.
Summers, R., and A. Heston. (1991). “The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950–1987.” *Quarterly Journal of Economics* 106(2), 1–41.