Statistical Hypothesis Testing for Returns to Scale Using Data Envelopment Analysis

M. Fukushige[a] **and I. Miyara**[b]

[a] Graduate School of Economics, Osaka University, Osaka 560-0043, Japan (mfuku@econ.osaka-u.ac.jp)

[b] Graduate School of Economics, Osaka University, Osaka 560-0043, Japan

(cg087mi@srv.econ.osaka-u.ac.jp)

Abstract: Data envelopment Analysis (DEA) is a kind of non-parametric and non-statistical tools for measuring technological efficiencies. Recently, several researchers adopt this method to evaluate the efficiencies of firms, public utilities, and so on. Because this method is not based on the statistical setup, we cannot test hypothesis where the production function should be assumed. In this paper, we propose several tests for returns to scale using different DEA models. We construct test statistics for the hypothesis about the returns to scale in production technology without any assumption on the true distribution of the technical inefficiencies. We conduct a simulation study on the size and power of the proposed test statistics under several conditions: Cobb-Douglas production function with half-normally and exponentially distributed error terms.

Keywords: *Data Envelopment Analysis; Test for returns to scale; Monte Carlo simulation*

1. INTRODUCTION

Data Envelopment Analysis (DEA) has been applied to evaluate the efficiencies in the economic activities like public or private service production. This is a kind of non-parametric and non-statistical tools, so the estimated inefficiencies depend on the assumption whether the production technology is constant returns to scale or not. From the researchers' or policy makers' point of view, whether the production technology is increasing or decreasing returns to the scale is one of the most important characteristics to make decisions. However, only few attempts have been made at the DEA-based statistical tests for the returns to scale. Banker (1996) surveys statistical tests using DEA but these tests are not enough to use in an empirical research because their asymptotic distribution and finite sample properties of them are not clear. Generally, DEA is a kind of non-parametric and non-statistical tools, so it is difficult to obtain their asymptotic distribution. Therefore, in the present paper, we propose some testing procedures and conduct a simulation study to investigate their finite sample properties.

The paper consists as follows. Section 2 introduces the new testing procedures for the returns to scale. Section 3 explains the setups of the Monte Carlo simulation. Section 4 presents simulation results. Section 5 is a concluding remark.

2. NEW TESTS FOR RETURNS TO SCALE USING DIFFERENT DEA MODELS

2.1. DEA models with different assumption of returns to scale

There are some models with different assumption in DEA; CCR, IRS, DRS, and BCC. (See Cooper, Seiford and Tone (2000).) CCR model is proposed by Charnes, Cooper and Rhodes (1978) and is assumed the frontier to be constant returns to scale. The input oriented CCR model is written as the following linear programming problem:

$$\min_{\lambda} \quad \theta_i$$

$$s.t. \quad y_{ij} \le \sum_{j=1}^{J} \lambda_j y_{jm}$$

$$\sum_{j=1}^{J} \lambda_j y_{jm} \le \theta_i x_{jn} \tag{1}$$

$$\lambda_j \ge 0$$

$$m = 1, \dots, M$$

$$n = 1, \dots, N$$

$$j = 1, \dots, J$$

where j=1,…,J (number of units), m=1,…,M (number of outputs), n=1,…,N (number of inputs).

IRS, DRS and BCC models have slightly different assumptions in Equation 1; they include a constraint on the multiplier, λ_i. IRS model assumes that the frontier exhibits increasing returns to scale: $\sum_{j=1}^{J} \lambda_j \geq 1$. DRS model assumes the frontier to be decreasing returns to scale: $\sum_{j=1}^{J} \lambda_j \leq 1$. BCC model, developed by Banker, Charnes and Cooper (1984), assumes the frontier to be variable returns to scale: $\sum_{j=1}^{J} \lambda_j = 1$.

2.2. Relationships between the true production function and the efficiency scores in DEA models

We often obtain different efficiency scores by applying different DEA models because there exists the difference between the returns to scale in the true production technology and those assumed by the DEA model that we applied. In the present paper, to simplify the explanation, we figure the relationships between the true production technology and the estimated efficiency scores with one input and one output case as an example.

Firstly, let us see the case that the true frontier production function, $y = f(x)$, is increasing returns to scale. See Figure 1. All points from A to F represent productions. Point A, B, C and D are on the true frontier function. The estimated frontiers by DEA models are as follows. The estimated frontier by CCR model is a straight line Ol passing through point D, the DRS model frontier is ODd, the IRS model frontier is aAl_1, and the BCC model frontier is a linear envelop, aADd. The efficiency score of each point varies according to the DEA models. For example, the efficiency score of point B estimated by CCR or that estimated by DRS expresses ratio of Bb_1 to b_1b_2, that estimated by IRS or BCC model is ratio of Bb to bb. The former is lower than the latter. Then, let $\overline{\theta}^i$ be the mean of the estimated efficiencies from model i (for i=CCR, IRS, DRS, BCC), the following inequality tends to hold;

$$\overline{\theta}^{BCC} \geq \overline{\theta}^{IRS} > \overline{\theta}^{DRS} \geq \overline{\theta}^{CCR}. \qquad (2)$$

The mean of BCC scores is the highest and the mean of CCR is the lowest among all of means because of the inefficient production such as point A or F. Inequality between the mean of IRS scores and that of DRS scores holds except in some cases; there are some units with large inefficiencies which are improved in DRS model.

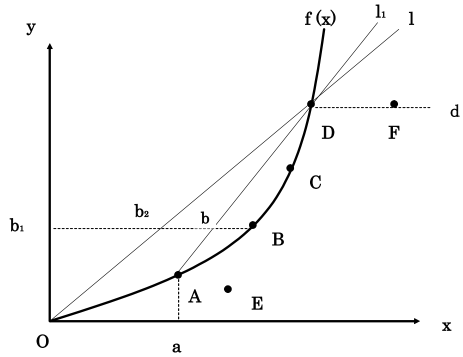

Figure 1. Increasing returns to scale technology and DEA models.

Secondly, let us see the case that the true frontier production function is decreasing returns to scale as Figure 2. The estimated frontiers in DEA models are as follows. The CCR model frontier is a straight line Ol passing through point A, the IRS model frontier is aAl, the DRS model frontier is OABCDd, and the BCC model frontier is a linear envelop, aOABCDd. Among the means, the following inequality tends to hold;

$$\overline{\theta}^{BCC} \geq \overline{\theta}^{DRS} > \overline{\theta}^{IRS} \geq \overline{\theta}^{CCR}. \qquad (3)$$

It is same as increasing returns to scale frontier that the highest mean is the BCC mean and the lowest is the CCR mean. However, inequality between the IRS and DRS means is different; the DRS mean is higher than the IRS mean except in some circumstances.

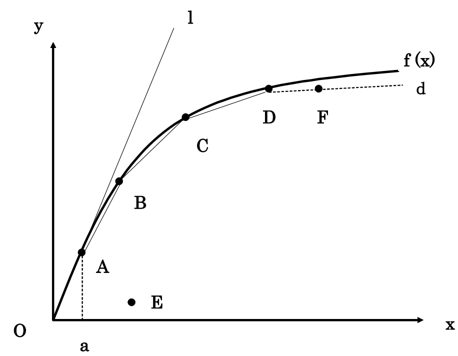

Figure 2. Decreasing returns to scale technology and DEA models.

Lastly, let us see the case that the true frontier function is constant returns to scale. In this case, the following inequality tends to hold;

$$\overline{\theta}^{BCC} \geq (\overline{\theta}^{IRS}\ ,\ \overline{\theta}^{DRS}) \geq \overline{\theta}^{CCR}. \qquad (4)$$

If all production units are efficient, all of the means are same and equal to one. When there are inefficient units, the highest and the lowest mean are same as other cases: the mean of BCC scores is highest and the mean of CCR scores is lowest and the relationship between the IRS and DRS means is not clear. But, we can expect all the means take closer values in case of constant returns to scale.

In summary, we can see the following points from the examples above. Firstly, it holds always that the BCC mean is the highest and the CCR mean is the lowest. Secondly, the IRS mean is higher than the DRS mean in case of increasing returns to scale except some circumstances. Thirdly, the DRS mean is larger than the IRS mean. Finally, all the means take closer value in case of constant returns to scale.

2.3. New tests for returns to scale using DEA

We propose some testing procedures for returns to scale using relationship among the means of scores estimated by different models. We will test the equality of $\overline{\theta}^{IRS} = \overline{\theta}^{DRS}$ and $\overline{\theta}^{BCC} = \overline{\theta}^{CCR}$ using the test of equality of means or sign test. The reason why we adopt sign test is that means can be easily affected by outliers. When the alternative hypothesis that $\overline{\theta}^{IRS} > \overline{\theta}^{DRS}$ is accepted, the true frontier function can be regard as increasing returns to scale. When the hypothesis that $\overline{\theta}^{DRS} > \overline{\theta}^{IRS}$ is accepted, the true frontier function can be regard as decreasing returns to scale. When the hypothesis that $\overline{\theta}^{IRS} = \overline{\theta}^{DRS}$ or $\overline{\theta}^{BCC} = \overline{\theta}^{CCR}$ is not rejected, the true frontier function can be regard as constant returns to scale. We suggest the following new four testing procedures.

A) Test of equality of mean of IRS scores and that of DRS scores.

・Case of $\overline{\theta}^{IRS} > \overline{\theta}^{DRS}$:

H_0: $\overline{\theta}^{IRS} = \overline{\theta}^{DRS}$ vs. H_1: $\overline{\theta}^{IRS} > \overline{\theta}^{DRS}$.

The test statistics is:

$$z_1 = \frac{\Delta}{\sqrt{\frac{S_{IRS}^2}{n} + \frac{S_{DRS}^2}{n}}} \tag{5}$$

where $\Delta \equiv \overline{\theta}^{IRS} - \overline{\theta}^{DRS}$, S stands for standard deviation and n is the number of production units.

Reject the null hypothesis when the test statistics is larger than 1.645, the upper 5% critical value from standard normal distribution. Then accept the alternative hypothesis (increasing returns to scale).

・Case that $\overline{\theta}^{DRS} > \overline{\theta}^{IRS}$:

H_0: $\overline{\theta}^{IRS} = \overline{\theta}^{DRS}$ vs. H_1: $\overline{\theta}^{DRS} > \overline{\theta}^{IRS}$.

Δ in Equation 5 is replaced with $\Delta \equiv \overline{\theta}^{DRS} - \overline{\theta}^{IRS}$.

B) Sign test of the probability that number of units, whose scores are improved in IRS than DRS, is more than half of all units

・Case of $\#(\theta_i^{IRS} > \theta_i^{DRS}) > n/2$:

H_0: $P(\theta_i^{IRS} > \theta_i^{DRS}) = 0.5$

H_1: $P(\theta_i^{IRS} > \theta_i^{DRS}) > 0.5$.

The test statistics is:

$$z_2 = \frac{X - (n/2)}{\sqrt{n/4}} \tag{6}$$

where X is the number of units whose IRS score is larger than DRS score, n is the number of units.

Reject the null hypothesis when the test statistics is larger than 1.645, the upper 5% critical value from standard normal distribution.

・Case that $\overline{\theta}^{DRS} > \overline{\theta}^{IRS}$:

H_0: $P(\theta_i^{DRS} > \theta_i^{IRS}) = 0.5$

H_1: $P(\theta_i^{DRS} > \theta_i^{IRS}) > 0.5$

and X is replaced with the number of units whose DRS score is larger than IRS score.

C) Test of equality of the BCC score mean and the CCR score mean and the equality of the IRS scores mean and the DRS scores mean.

1st step: Test the equality of the mean of BCC scores and CCR scores

H_0: $\overline{\theta}^{BCC} = \overline{\theta}^{CCR}$ vs. H_1: $\overline{\theta}^{BCC} > \overline{\theta}^{CCR}$.

The test statistics is:

$$z_3 = \frac{\overline{\theta}^{BCC} - \overline{\theta}^{CCR}}{\sqrt{\frac{S_{BCC}^2}{n} + \frac{S_{CCR}^2}{n}}} \tag{7}$$

Reject the null hypothesis when the test statistics is larger than 1.645, the upper 5% critical value from standard normal distribution, then proceed next step. If not rejected, accept the null hypothesis (constant returns to scale).

2nd step: Test the equality of the mean of IRS scores and DRS scores. It is same as (A).

D) Test of equality of the BCC score mean and the CCR score mean and sign test of the probability that number of units, whose scores are improved in IRS than DRS, is more than half of all units.

1st step: This step is same as first step of (C). When the null hypothesis is rejected, then proceed to next step. If not rejected, accept the null hypothesis (constant returns to scale).

2nd step: This step is same as (B).

3. SETTING FOR MONTE CARLO SIMULATIONS

We conduct Monte Carlo simulation to study the finite sample properties of the proposed tests.

3.1. Production Technology

Generally, the DEA can be applied to multiple inputs and multiple output production technology. But, for its simplicity, according to Banker (1996), we specify the following Cobb-Douglas production technologies;

Homogenous production function:

$$Q = 10x_1^{0.6}x_2^{0.4} \tag{8a}$$

Concave production function:

$$Q = 10x_1^{0.4}x_2^{0.3} \tag{8b}$$

Convex production function:

$$Q = 10x_1^{0.8}x_2^{0.5} \tag{8c}$$

where x_1 and x_2 are each drawn randomly and independently from uniform probability distribution over the interval [5, 15]. The first technology exhibits constant returns to scale for all input value. The second production function is concave, exhibiting decreasing returns to scale for all inputs. The third production function is convex, exhibiting increasing returns to scale for all inputs. These settings are same as Banker (1996) adopted.

3.2. Inefficiency Distributions

The distribution of the inefficiency θ_j for any observation j must be above 1, we write $\theta_j = 1 + \varphi_j$, $\varphi_j \geq 0$. We consider four different distributions as follows:

$$\varphi_j \sim |N(0,0.20)| \tag{9}$$

$$\varphi_j \sim |N(0,0.063)| \tag{10}$$

$$\varphi_j \sim EXP(0.159) \tag{11}$$

$$\varphi_j \sim EXP(0.05) \tag{12}$$

$EXP(\mu)$ represents the exponential distribution with mean μ. The random variable of (9) and (10) is generated from half normal distribution. Both of means of (9) and (11) are same, 0.159. Also both of the means of (10) and (12) are same, 0.05. The distribution of (9) and (11) are same as Banker (1996) adopted.

3.3. Sample Size and Simulated Observation

As noted above, the value of inputs vector for each observation j are generated from a uniform distribution. The corresponding value of the output for each observation j, Y_j, is obtained using the production function (8) and the simulated value of the inefficiency term θ_j. We generate Y_j as follows:

$$Y_j = Q_j / \theta_j = f(X_j)/\theta_j \tag{13}$$

We consider the case that the sample size is 100. Totally we consider 12 setups: 3 production technologies, 4 inefficiency distributions and 1 sample size. We replicate 500 drawings for each setup. We use GAUSS 5.0 to generate random variables, solve linear programming for the DEA estimates, and construct test statistics.

4. RESULTS OF SIMULATIONS

In this paper, we compare the proposed tests above and the tests introduced by Banker (1996). Before the discussion of simulations results, we will explain the tests for returns to scale using DEA that are introduced by Banker (1996).

E) Test of goodness of fit

- $$t_4 = \frac{\sum_{j=1}^{N}(\theta_j^{CCR}-1)^2}{\sum_{j=1}^{N}(\theta_j^{BCC}-1)^2}$$

which follows the F distribution with (N,N) if the distribution of θ_j^{CCR} and θ_j^{BCC} are same. If not rejected, accept the constant returns to scale hypothesis.

F) Kolmogorov-Smlirnov Test

- $$\max\{F(\theta_j^{CCR}) - F(\theta_j^{BCC}) \mid j = 1,\ldots,N\}$$

where F(.) is the empirical distribution function. If not rejected, accept the constant returns to scale hypothesis.

Table 1: Summary of statistical test results (sample size=100)

Panel A:True production technology is constant return to scale

Type of Tests		Hypothesis		Inefficiency distribution			
				Exponetial		Half Normal	
		Null	Alt.	$\mu=0.05$	$\mu=0.157$	$\mu=0.05$	$\mu=0.157$
A	$\theta^{IRS} > \theta^{DRS}$	CRS	IRS	0 (0)	0.028 (0)	0.518 (0.238)	1 (0.948)
	$\theta^{DRS} > \theta^{IRS}$	CRS	DRS	0 (0)	0 (0)	0 (0)	0 (0)
B	$\theta^{IRS} > \theta^{DRS}$	CRS	IRS	0.014 (0.010)	0.232 (0.170)	0.59 (0.556)	0.894 (0.876)
	$\theta^{DRS} > \theta^{IRS}$	CRS	DRS	0.598 (0.516)	0.232 (0.182)	0.068 (0.044)	0.008 (0.006)
C	$\theta^{BCC} = \theta^{CCR}$	CRS	IRS, DRS	0.120 (0.028)	0.166 (0.038)	0.834 (0.526)	1 (1)
	$\theta^{IRS} > \theta^{DRS}$	CRS	IRS	0 (0)	0.028 (0)	0.520 (0.24)	1 (0.948)
	$\theta^{DRS} > \theta^{IRS}$	CRS	DRS	0 (0)	0 (0)	0 (0)	0 (0)
D	$\theta^{BCC} = \theta^{CCR}$	CRS	IRS, DRS	0.120 (0.028)	0.166 (0.038)	0.834 (0.526)	1 (1)
	$\theta^{IRS} > \theta^{DRS}$	CRS	IRS	0.004 (0.002)	0.052 (0.012)	0.524 (0.340)	0.894 (0.876)
	$\theta^{DRS} > \theta^{IRS}$	CRS	DRS	0.060 (0.010)	0.024 (0.002)	0.048 (0.012)	0.008 (0.006)
E	$\theta^{BCC} = \theta^{CCR}$	CRS	IRS,DRS	0.054 (0.006)	0.100 (0.010)	0.286 (0.014)	0.936 (0.504)
	$\theta^{IRS} = \theta^{DRS}$	CRS	IRS,DRS	0 (0)	0 (0)	0 (0)	0 (0)
F	$\theta^{BCC} = \theta^{CCR}$	CRS	IRS,DRS	1 (1)	1 (1)	1 (1)	1 (1)
	$\theta^{IRS} = \theta^{DRS}$	CRS	IRS,DRS	1 (1)	1 (1)	1 (1)	1 (1)

Panel B:True production technology is decreasing returns to scale

Type of Tests		Hypothesis		Inefficiency distribution			
				Exponetial		Half Normal	
		Null	Alt.	$\mu=0.05$	$\mu=0.157$	$\mu=0.05$	$\mu=0.157$
A	$\theta^{IRS} > \theta^{DRS}$	CRS	IRS	0 (0)	0 (0)	0 (0)	0.440 (0.204)
	$\theta^{DRS} > \theta^{IRS}$	CRS	DRS	1 (1)	1 (1)	0.632 (0.420)	0 (0)
B	$\theta^{IRS} > \theta^{DRS}$	CRS	IRS	0 (0)	0 (0)	0 (0)	0.258 (0.208)
	$\theta^{DRS} > \theta^{IRS}$	CRS	DRS	1 (1)	1 (1)	0.976 (0.960)	0.142 (0.082)
C	$\theta^{BCC} = \theta^{CCR}$	CRS	IRS, DRS	1 (1)	1 (1)	1 (1)	1 (0.998)
	$\theta^{IRS} > \theta^{DRS}$	CRS	IRS	0 (0)	0 (0)	0 (0)	0.442 (0.206)
	$\theta^{DRS} > \theta^{IRS}$	CRS	DRS	0.998 (0.998)	0.998 (0.998)	0.630 (0.420)	0 (0)
D	$\theta^{BCC} = \theta^{CCR}$	CRS	IRS, DRS	1 (1)	1 (1)	1 (1)	1 (0.998)
	$\theta^{IRS} > \theta^{DRS}$	CRS	IRS	0 (0)	0 (0)	0 (0)	0.260 (0.208)
	$\theta^{DRS} > \theta^{IRS}$	CRS	DRS	0.998 (0.998)	0.998 (0.998)	0.974 (0.958)	0.142 (0.082)
E	$\theta^{BCC} = \theta^{CCR}$	CRS	IRS,DRS	1 (1)	1 (1)	0.912 (0.504)	0.880 (0.270)
	$\theta^{IRS} = \theta^{DRS}$	CRS	IRS,DRS	1 (1)	1 (1)	0.136 (0.004)	0 (0)
F	$\theta^{BCC} = \theta^{CCR}$	CRS	IRS,DRS	1 (1)	1 (1)	1 (1)	1 (1)
	$\theta^{IRS} = \theta^{DRS}$	CRS	IRS,DRS	1 (1)	1 (1)	1 (1)	1 (1)

Panel C:True production technology is increasing returns to scale

Type of Tests		Hypothesis		Inefficiency distribution			
				Exponetial		Half Normal	
		Null	Alt.	$\mu=0.05$	$\mu=0.157$	$\mu=0.05$	$\mu=0.157$
A	$\theta^{IRS} > \theta^{DRS}$	CRS	IRS	1 (1)	1 (1)	1 (1)	1 (1)
	$\theta^{DRS} > \theta^{IRS}$	CRS	DRS	0 (0)	0 (0)	0 (0)	0 (0)
B	$\theta^{IRS} > \theta^{DRS}$	CRS	IRS	1 (1)	1 (1)	0.998 (0.998)	0.998 (0.994)
	$\theta^{DRS} > \theta^{IRS}$	CRS	DRS	0 (0)	0 (0)	0 (0)	0 (0)
C	$\theta^{BCC} = \theta^{CCR}$	CRS	IRS, DRS	1 (1)	1 (1)	1 (1)	1 (1)
	$\theta^{IRS} > \theta^{DRS}$	CRS	IRS	1 (1)	1 (1)	1 (1)	1 (1)
	$\theta^{DRS} > \theta^{IRS}$	CRS	DRS	0 (0)	0 (0)	0 (0)	0 (0)
D	$\theta^{BCC} = \theta^{CCR}$	CRS	IRS, DRS	1 (1)	1 (1)	1 (1)	1 (1)
	$\theta^{IRS} > \theta^{DRS}$	CRS	IRS	1 (1)	1 (1)	0.998 (0.998)	0.998 (0.994)
	$\theta^{DRS} > \theta^{IRS}$	CRS	DRS	0 (0)	0 (0)	0 (0)	0 (0)
E	$\theta^{BCC} = \theta^{CCR}$	CRS	IRS,DRS	1 (1)	1 (1)	1 (1)	1 (1)
	$\theta^{IRS} = \theta^{DRS}$	CRS	IRS,DRS	0 (0)	0 (0)	0 (0)	0 (0)
F	$\theta^{BCC} = \theta^{CCR}$	CRS	IRS,DRS	1 (1)	1 (1)	1 (1)	1 (1)
	$\theta^{IRS} = \theta^{DRS}$	CRS	IRS,DRS	1 (1)	1 (1)	1 (1)	1 (1)

The number reported in each cell is the parcentage (out of 500 itarations) for which the coresponding test statistics is rejected at the 5% significant level. The correspond number in parentheses in each cell indicates the percentage for which the test statistic is rejected at the 1% significant level.

Table 1 presents the results of six kinds of tests for returns to scale. The percentage in each cell represents the percentage of that the null hypothesis is rejected. From Panel 1, size of the test (E) introduced in Banker is 0% regardless of the inefficiency distribution. On the other hand, size of test (F) is 100%. Size of the proposed test (A), (C) and (D) are less than 6% in the case of exponential distributed inefficiency.

When the true production technology is decreasing or increasing returns to scale, the percentage of rejection represents the power of test. As for the decreasing returns to scale, the power of test (F) is 100% in all cases. The power of test (E) is over 85% regardless of distribution. The power of test (B) and (D) are over 95% except the case of half normal distribution with larger mean. As for the case of exponential distribution, the power of test (B) and (D) are over 99%. Half normal distribution has larger variance than exponential distribution with same mean as half normal distribution. Therefore one of reasons why the power of test using half normal distributed inefficiency is lower may be occurrence of outliers.

As for the case of increasing returns to scale technology, the power of all tests is over 99% regardless of distribution in contrast with decreasing returns to scale technology. This result may be brought by that input value range is narrow. For increasing returns to scale technology, the difference between the efficiency of IRS and DRS is larger as input value is larger (See Figure1).

5. CONCLUSIONS

In this paper, we propose new tests for returns to scale with some different DEA models. To investigate the size and powers of the test statistics, simulations are conducted under several conditions: Cobb-Douglas with exponential and half normal distributed error.

From results of simulations, test (B), (C) and (E) have good property in the case of the exponential distributed inefficiency. However, there is no reason that the inefficiency follows only exponential distribution. One of reasons that the test statistics do not work well with the half normal distributed inefficiency can be effects of outlier for the mean. We have to pay attention to use these tests in case of inefficiency distributed half normal.

6. ACKNOWLEDGEMENT

This research was supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aids for Scientific Research (C), 13630034.

7. REFERENCES

Banker, Rajiv D., Hypothesis Tests Using Data Envelopment Analysis, *Journal of Productivity Analysis,* 7, 139-159, 1996.

Banker, Rajiv D., Charnes, A. and Cooper, W. W., Some models for estimating technical and scale inefficiencies in data envelopment analysis, *Management Science*, 30, 1078-1092, 1984.

Charnes, A., Cooper, W. W. and Rhodes, E, Measuring the efficiency of decision making units, *European Journal of Operational Research*, 2, 213-222, 1978.

Cooper, William W., Seiford, Lawrence M. and Tone, Kaoru, Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software, 2000.