Frame Work of Data Envelopment Analysis—A Model to Evaluate the Environmental Efficiency of China's Industrial Sectors

TAO ZHANG[#,1]

[#]*Department of Economics, National University of Ireland, Galway, Ireland*

Objective To evaluate the environmental and technical efficiencies of China's industrial sectors and provide appropriate advice for policy makers in the context of rapid economic growth and concurrent serious environmental damages caused by industrial pollutants. **Methods** A data of envelopment analysis (DEA) framework crediting both reduction of pollution outputs and expansion of good outputs was designed as a model to compute environmental efficiency of China's regional industrial systems. **Results** As shown by the geometric mean of environmental efficiency, if other inputs were made constant and good outputs were not to be improved, the air pollution outputs would have the potential to be decreased by about 60% in the whole China. **Conclusion** Both environmental and technical efficiencies have the potential to be greatly improved in China, which may provide some advice for policy-makers.

Key words: Technical efficiency; Environmental efficiency; Directional distance function; Technical-environmental efficiency; Data of envelopment analysis; China

INTRODUCTION

China is now experiencing rapid economic growth by developing industrial production. However, extremely rapid output growth has been accompanied with serious environmental damage and has posed in turn a threat to public health[1-3]. As a result, many cities in China are among the worst polluted urban areas in the world and are faced with serious public health hazards associated with environmental pollution[4-5]. Thus, in China, a measurement of the environmental performance of industrial manufacturing sectors is essential to identify the industrial sources of pollution, which is necessary for devising efficient control strategies. However, most of literature on regional environmental assessment of China only focuses on pollution indice or the impact of environmental regulations[6-7]. Zhang and Xue[8-9] analyzed the environmental efficiency of China's agricultural sectors producing vegetables and crops, while ignoring the industrial system. Consequently, few studies have empirically measured the environmental efficiency of China's industrial sectors at a nation-wide level and the perspectives of productivity efficiency.

The measurement of productivity has traditionally focused on measuring the outputs of production units relative to inputs of production. The productivity of an individual production unit could be measured in terms of its ability to minimize input usage while producing given outputs, or to maximize output production with given inputs. In other words, any firm at full efficiency should be operated at maximum potential output levels, and any deviation from the frontier would be used to measure its inefficiency. Presently, the most popular methods used in efficiency evaluation involve non-parametric and mathematical programming framework, as well as stochastic and econometric framework. Since data of envelopment analysis (DEA) can be applied in multi-output situations and eliminate the need for parametric assumption of the underlying technology, DEA is a widely-used linear programming technique for conducting such an evaluation.

However, with increasing interest in assessing environmental performance of industrial sectors, the above traditional method, typically ignoring the production of by-products such as pollution, fails to provide information on assessing environmental performance. Recently, some studies have partially satisfied the need of assessing environmental performance by incorporating the environmental

[1]Correspondence should be addressed to Tao ZHANG. Department of Economics, National University of Ireland, Galway, Ireland. Tel: 353-87-6777215. E-mail: taozhang7608@hotmail.com

Biographical note of the first author: Tao ZHANG, male, born in 1976, Walsh research fellow at Teagasc, majoring in environmental economics and sustainable development.

0895-3988/2009
CN 11-2816/Q
Copyright © 2009 by China CDC

factor into adjusted measures of productivity or constructing environmental index. Being different from traditional studies of productivity ignoring the reduced emissions of pollutants, these studies implicitly assume that inputs into the pollution abatement activities are productive.

Environmental performance assessment can be defined as an analysis to study the environmental performance of a given firm in different time periods, or to compare the environmental performance of different firms in a single time period by using some analytical tools. While traditional DEA models account for only two categories of variables (inputs and outputs), their use in environmental performance assessment should consider three kinds of variables, namely, inputs, desirable and undesirable (environmentally detrimental or pollution) outputs[10]. Here, the practical difficulty in incorporating the environmental factor into adjusted measures of productivity is the task of assigning weights to the bad outputs.

Although some DEA frameworks have been used for environmental performance assessment in the past decades[11-12], most of them are only able to supply undesirable (environmentally detrimental or pollution) output-oriented technical efficiency (environmental efficiency) or directional distance function in which the good and bad outputs are treated asymmetrically.

The undesirable output-oriented technical efficiency cannot provide overall productivity measurement. In addition, although the directional distance function credits reduction of environmentally detrimental outputs and expansion of good outputs, it cannot supply the same efficiency score for either the reduction of environmentally detrimental outputs or the expansion of good outputs at the same time. In other words, the method of directional distance function can provide both the technical efficiency measuring the expansion of good outputs and the environmental efficiency measuring the reduction of pollutions simultaneously, but their scores are different. In addition, scores of the calculated technical efficiency in the directional distance function model only can range from 0.5 to 1.

In this paper, we established a DEA framework, which not only can credit the reduction of pollution outputs and the expansion of good outputs simultaneously but also can calculate the same values for both technical and environmental efficiencies.

METHODS

To describe the DEA framework used in this paper, we started from observations of K units that use N inputs $x = (x_1, ..., x_N) \in R_+^N$ to produce M desirable outputs $y = (y_1, ..., y_M) \in R_+^M$ while releasing W undesirable outputs (environmentally detrimental outputs). To account for the characteristics of production process producing negative externalities, a distinctive technology set should be specified. As a definition of environmental efficiency (referred to as EE in this paper.), we used the following reference technology S proposed by Färe *et al.*[13].

$$
\begin{aligned}
S = \{(x,y,b): & \sum_{k=1}^{K} z^k x_n^k \leqslant x_n, n = 1,...,N \\
& \sum_{k=1}^{K} z^k y_m^k \geqslant y_m, m = 1,...,M \\
& \sum_{k=1}^{K} z^k b_w^k = b_w, w = 1,...,M \\
& z^k \geqslant 0, k = 1,...,K\} \qquad (1)
\end{aligned}
$$

The first two sets of constraints indicate a vector of desirable (good) outputs which are produced by a vector of inputs x, based on the reference technology S. The third constraint indicates a vector of undesirable (pollution) outputs b. From the above function, it is manifested that the good outputs and inputs are freely disposable. Here, we assumed that it was costly to reduce the bad outputs by imposing that b and y were jointly weaky disposable. The weak disposability indicates that it is feasible to reduce good and bad outputs proportionally, while it is infeasible to reduce only the bad output without the consumption of resources that otherwise could have been used to produce the good output. According to Färe[14], another property the model holds is that desirable and undesirable outputs are null joint, which means that it is technically impossible to produce good outputs without simultaneously producing bad outputs. Based on the above three properties and the idea of specified technology set, an undesirable output-oriented efficiency index can be described as:

$$\varepsilon = \inf\{\theta : (x, y, \theta b) \in S\}. \qquad (2)$$

Function (2) can be directly treated as the definition of EE. Here, ε is environmental efficiency, which is equal to θ. If ε<1, a sample is not lying on the frontier of the production set, and an improvement in environmental performance is possible for this sample. If ε=1, it indicates that, in the current status of technology reflected by the frontier of the production set, no significant improvements could be made in the sample. However, while this index can measure the reduction of bad outputs, it ignores the measurement of expansion of good outputs. Obviously, function (2) can provide us with environmental efficiency, but cannot give us an overall productivity index.

A significant effort was made by Chung, Färe, and Grosskopf[15] who developed a directional distance function crediting a producer for simultaneously reducing the bad outputs and increasing the good outputs. They believed that the original Malmquist index using Shephard output distance functions to represent technology[16] could expand both the good and bad outputs proportionally as much as feasible. The framework of Shephard output distance function considering both the good and bad outputs can be defined as

$$D(x,y,b)^{-1} = \sup\{\theta : (x,\theta y,\theta b) \in S\}. \quad (3)$$

where technical efficiency can be computed using 1/θ (referred to as TE in this paper). Since both types of outputs are expanded at the same rate in Shephard output distance functions, it cannot be used to measure environmental performance. So, Chung, Färe, and Grosskopf have developed a new index that could explicitly credit firms or industries for reductions in undesirable outputs, while providing a measurement of "true" productivity. This method is based on the basic idea of the Luenberger productivity index which takes account of both input reductions and output improvements when measuring efficiency[17-18]. In their studies, the directional distance function seeking to increase the good outputs while simultaneously decreasing the bad outputs is defined as

$$\vec{D}_o(x,y,b;g) = \sup\left\{\theta : (y,b) + \theta g \in S\right\}. \quad (4)$$

where "g" is the vector of "directions" in which outputs are scaled, θ is the expansion of the desirable outputs and contraction of the undesirable outputs when the expansion and contraction are identical proportions for a given level of inputs. The above directional distance functions can be expressed as solutions to linear programming problems as follows:

$$\vec{D}_o(x,y,b;g) = Max\theta \quad (5)$$

$$s.t. \sum_{k=1}^{K} z^k x_n^k \leqslant x_n, n = 1,..., N,$$

$$\sum_{k=1}^{K} z^k y_m^k \geqslant (1+\theta) y_m, m = 1,..., M$$

$$\sum_{k=1}^{K} z^k b_w^k = (1-\theta) b_w, w = 1,..., W$$

$$z^k \geqslant 0, k = 1,..., K$$

Since the technical and environmental efficiencies in this model are calculated simultaneously using direct distance function, we call them DTE and DEE in this model. Here, according to the definition of technical efficiency and environmental efficiency, DTE can be computed as DTE=1/(1+θ), and DEE as DEE=1-θ. Thus, although the method of directional distance function can provide both the technical efficiency measuring the expansion of good outputs and the environmental efficiency measuring the reduction of bad outputs simultaneously, their scores are different.

In this paper, we took the advantages of both Farrell's output measure of technical efficiency[19] and Shephard's distance function to establish our DEA framework. Here, we also sought to increase the good outputs while decreasing the bad outputs simultaneously. The technical-environmental efficiency (referred to as TEE in this paper.) index can be defined as

$$TEE^{-1} = \sup\left\{\theta : (y\theta, x, b/\theta) \in S\right\}. \quad (6)$$

where TEE=1/θ or defined as

$$TEE = \inf\left\{\theta : (y/\theta, x, b\theta) \in S\right\}. \quad (7)$$

where TEE=θ. The above functions can expand the good outputs and contract the bad outputs as much as feasible at the same time. Obviously, function (6) is the reciprocal of function (7). Such a relationship also exists between the Farrell output function and Shephard's output distance function. The functions can be calculated as optimal solutions to the following programming problems

$$TEE^{-1} = Max\theta \quad (8)$$

$$s.t. \sum_{k=1}^{K} z^k x_n^k \leqslant x_n, n = 1,..., N,$$

$$\sum_{k=1}^{K} z^k y_m^k \geqslant \theta y_m, m = 1,..., M$$

$$\sum_{k=1}^{K} z^k b_w^k = b_w/\theta, w = 1,..., W$$

$$z^k \geqslant 0, k = 1,..., K$$

Or

$$TEE = Min\theta \quad (9)$$

$$s.t. \sum_{k=1}^{K} z^k x_n^k \leqslant x_n, n = 1,..., N,$$

$$\sum_{k=1}^{K} z^k y_m^k \geqslant y_m/\theta, m = 1,..., M$$

$$\sum_{k=1}^{K} z^k b_w^k = \theta b_w, w = 1,..., W$$

$$z^k \geqslant 0, k = 1,..., K$$

After the empirical experiment in which we used extremely small value as a good output, we found that the value of TEE approached zero as a limit. Because this method satisfies the basic idea of efficiency measurement, it is more applicable than directional distance function in efficiency analysis.

RESULTS

For the empirical measurement, we established a province-level database from official yearbooks available in China: China Environment Yearbooks (2005), China Statistical Yearbooks (2005), and China Industry Economy Statistical Yearbooks (2005). The term "province" refers to provinces, autonomous regions and municipalities directly under the central government. A data set in 2004 for the industrial sector in China's 30 provinces was established, which consists of provincial level observations on industrial outputs and inputs, as well as emissions of pollutants.

The two inputs considered were the provincial aggregates of industrial employment and capital. The data of one undesirable output (pollutant emission) were the emissions of air pollution by the industrial sector. We treated the total volume of waste gas emission from the industry as air pollution output.

The calculated values for the various efficiency indexes, which were mentioned in the former section, are listed in Table 1. In addition, the discrepancies between the different efficiency scores and rankings obtained from four programming models also can be found in Table 1.

TABLE 1

Results of DEA

Province	EE	TE	DTE	DEE	TEE
Beijing	0.492(9)	0.865(15)	0.777(12)	0.713(12)	0.758(15)
Tianjin	0.902(2)	0.820(19)	0.958(2)	0.956(2)	0.960(2)
Hebei	0.268(20)	0.970(2)	0.761(13)	0.686(13)	0.798(11)
Shanxi	0.180(24)	0.913(11)	0.598(23)	0.327(23)	0.555(23)
Inner Mongolia	0.151(26)	0.936(6)	0.576(26)	0.264(26)	0.510(26)
Liaoning	0.311(16)	0.941(5)	0.704(17)	0.579(17)	0.780(13)
Jilin	0.292(19)	0.843(17)	0.666(19)	0.498(19)	0.616(22)
Heilongjiang	0.347(14)	0.899(12)	0.736(15)	0.641(15)	0.742(16)
Shanghai	1.000(1)	1.000(1)	1.000(1)	1.000(1)	1.000(1)
Jiangsu	1.000(1)	1.000(1)	1.000(1)	1.000(1)	1.000(1)
Zhejiang	1.000(1)	1.000(1)	1.000(1)	1.000(1)	1.000(1)
Anhui	0.528(7)	1.000(1)	1.000(1)	1.000(1)	1.000(1)
Fujian	0.711(3)	0.756(21)	0.878(3)	0.861(3)	0.872(4)
Jiangxi	0.308(17)	0.969(3)	0.823(8)	0.786(8)	0.827(8)
Shandong	0.571(6)	0.921(9)	0.876(4)	0.859(4)	0.877(3)
Henan	0.339(15)	0.962(4)	0.857(5)	0.833(5)	0.859(5)
Hubei	0.496(8)	0.890(13)	0.822(9)	0.784(9)	0.820(10)
Hunan	0.349(13)	0.933(7)	0.836(7)	0.805(7)	0.839(6)
Guangdong	1.000(1)	1.000(1)	1.000(1)	1.000(1)	1.000(1)
Guangxi	0.362(11)	0.920(10)	0.756(14)	0.678(14)	0.771(14)
Hainan	0.625(5)	0.547(23)	0.813(11)	0.769(11)	0.791(12)
Sichuan	0.354(12)	0.825(18)	0.725(16)	0.620(16)	0.717(17)
Guizhou	0.155(25)	1.000(1)	0.588(24)	0.300(24)	0.646(19)
Yunnan	0.464(10)	0.926(8)	0.815(10)	0.773(10)	0.825(9)
Tibet	0.692(4)	0.193(24)	0.846(6)	0.818(6)	0.833(7)
Shaanxi	0.298(18)	0.785(20)	0.668(18)	0.502(18)	0.622(20)
Gansu	0.216(23)	0.868(14)	0.617(22)	0.379(22)	0.547(24)
Qinghai	0.237(22)	0.677(22)	0.625(21)	0.400(21)	0.519(25)
Ningxia	0.146(27)	1.000(1)	0.583(25)	0.286(25)	0.666(18)
Xinjiang	0.259(21)	0.864(16)	0.655(20)	0.474(20)	0.621(21)
Geometric Mean	0.397	0.846	0.773	0.639	0.764

The geometric mean of all EE scores was 0.397, indicating that, if we make other inputs constant and do not want to improve the good outputs, the bad outputs can probably be decreased by about 60% throughout China. The geometric mean of all TE scores was 0.846, indicating that, if we make all inputs constant and do not want to decrease the pollutant emissions, both the good and bad outputs have the potential to be increased by about 15% in the whole China. The geometric mean of DTE and DEE scores were 0.773 and 0.639 respectively, indicating that we can improve the good outputs by 23% and reduce the pollutant emissions by 36% simultaneously while making inputs constant. The geometric mean of TEE score was 0.764, suggesting that we can increase the good outputs by 24% and reduce the pollutant emissions by 24% simultaneously while making inputs constant. The estimated results in Table 1 show that both the environmental and technical efficiencies have the potential to be greatly improved.

We tested the differences in the indexes for the observations by using Anova test and Kruskal-Wallis test. Anova test is based on a single-factor (between-subjects) analysis of variance (ANOVA). The basic idea is that if the subgroups have the same mean, then the variability between the sample means (between groups) should be the same as the variability within any subgroup (within group). Kruskal-Wallis test is a median equality test for more than two subgroups. The basic idea is to rank the series from the smallest to the largest value, and to compare the sum of ranks from subgroup 1 to that from subgroup 2. Then, the null hypothesis of the same median is tested. If the groups have the same median, the values should be similar. The test results are shown in Table 2.

TABLE 2

Non-parametric Tests of Equality of the Indice

Included Indice	Anova (F)	Prob.>F	Kruskal-Wallis (χ^2)	Prob.> χ^2
TEE, DTE	0.028	0.869	0.007	0.935
TEE, DEE	3.239	0.077	2.1201	0.145
TEE, TE	5.360	0.024	8.099	0.004
TEE, EE	28.890	0.000	17.75	0.000
TEE, DTE, DEE, TE, EE	17.669	0.000	37.670	0.000

From the results in Table 2, we could see that there were significant differences between TEE and TE/ EE scores. Nevertheless, the hypothesis that TEE, DTE and DEE are the same and cannot be rejected, indicating that there are no significant differences between the scores of TEE, DTE and DEE, suggesting that, if there does not exist the observation in which inputs and bad outputs are high enough, and good outputs are low enough to produce a very low TEE score (lower than 0.5), the results of TEE and DTE/DEE models are similar.

However, although the geometric means of DTE and TEE were not significantly different, the discrepancies between the individual ranks of DTE and TEE were different. Only about 1/3 of all provinces held the same ranks for both DTE and TEE. In addition, the lowest score of DTE was 0.576, whereas the lowest score of TEE was 0.510, indicating that the limitation of DTE score is stronger when its score approaches 0.5.

DISCUSSION

From the estimated results, it is obvious that three provinces and one city (Jiangsu, Zhejiang, Guangdong, and Shanghai) holding the highest efficiency scores for all indicators (EE, TE, DTE, DEE, and TEE) are located in best-developed regions of China. These three provinces and one city contribute heavily to China's overall efficiency (geometric-mean in Table 1) in industries. In these three provinces and one city, although the absolute quantities of overall pollutant emissions are relatively higher than those in the other regions, the efficiency in these areas is the highest, indicating that industrial manufacturers in these areas can produce more outputs while yielding less pollutant emissions. The possible reason for the above situation is the different technology levels. In Jiangsu, Shanghai, Zhejiang and Guangdong, the economy and industries are relatively well developed, and therefore manufacturers in these regions have more financial sources to develop or obtain better technologies and management in production. The higher level of technology and management in these regions can improve the efficiency in production.

Although we measured the efficiency of China's industries by taking into account the productivity and air-pollutant emissions, it cannot provide more information on the factors which may potentially influence the efficiency estimates because DEA is not a stochastic model. A second-stage regression is a possible method but it is probably sensitive to DEA results, which should be reassessed in future.

In conclusion, we compared the original EE and TE with DEE and DTE, and accordingly introduced a new DEA framework to calculate the TEE that credits the reduction of undesirable outputs while simultaneously crediting increases in desirable outputs. An empirical example of China's industrial production in 2004 was provided by showing how to compute the indice using programming problems. The geometric mean of environmental efficiency shows that, if we make other inputs constant and do not want to improve the good outputs, the pollution outputs have the potential to be decreased by about 60% in the whole China. Both environmental and technical efficiencies can be greatly improved in China.

REFERENCES

1. Alebic-Juretic A, Frkovic A, Simic D (2001). The effect of air pollution on the outcome of pregnancies. *International Journal of Gynecology and Obstetrics* **75**(3), 315-316.
2. Bascom R, Bromberg P A, Costa D A, *et al*. (1996). Health effects of outdoor air pollution. *American Journal of Respiratory and Critical Care Medicine* **153**(1), 3-50.
3. Brunekreef B, Holgate S T (2002). Air pollution and health, LANCET **360**(9341), 1233-1242.
4. Zhang Y H, Chen C H, Chen G H, *et al*. (2006). Application of DALYs in measuring health effect of ambient air pollution: A case study in Shanghai, China. *Biomedical and Environmental Sciences* **19**(4), 268-272.
5. Jiang L L, Zhang Y H, Song G X, *et al*. (2007). A time series analysis of outdoor air pollution and preterm birth in Shanghai, China. *Biomedical and Environmental Sciences* **20**(5), 426-431.
6. Wang H, Jun B, David W, *et al*. (2004). Environmental performance rating and disclosure: China's GreenWatch program. *Journal of Environmental Management* **71**, 123-133.
7. Susmita D, LaplanteB, Mamingi N, *et al*. (2001). Inspections, pollution prices, and environmental performance: evidence from China. *Ecological Economics* **36**, 487-498.
8. Zhang T, Xue B D (2005). Environmental efficiency analysis of China's vegetable production. *Biomedical and Environmental Sciences* **18**(1), 21-30.
9. Zhang T (2008). Environmental performance in China's agricultural sector: A case study in corn production. *Applied Economics Letters* **15**(8), 641-645.
10. Tyteca D (1997). Linear programming models for the measurement of environmental performance of firms—concepts and empirical results. *Journal of Productivity Analysis* **8**, 183-197.
11. Zaim O (2004). Measuring environmental performance of state manufacturing through changes in pollution intensities: a DEA framework. *Ecological Economics* **48**(1), 37-47.
12. Färe R, Grosskopf S, Pasurka C (2001). Accounting for air pollution in measures of state manufacturing growth. *Journal of Regional Science* **41**, 381-409.
13. Färe R, Grosskopf S, Lovell C A K, *et al*. (1989). Multilateral productivity comparisons when some outputs are undesirable. *Review of Economics and Statistics* **71**, 90-98.
14. Färe R, Grosskopf S, Sancho H (2004). Environmental Performance: An Index Number Approach. *Resource and Energy Economics* **26**, 343-352.
15. Chung Y H, Färe R, Grosskopf S (1997). Productivity and Undesirable Outputs: A Directional Distance Function Approach. *Journal of Environmental Management* **51**, 229-240.
16. Shephard R W (1970). Theory of Cost and Production. Princeton: Princeton University Press.
17. Luenberger D G (1992a). Benefit functions and duality. *Journal of Mathematical Economics* **21**, 461-486.
18. Luenberger D G (1992b). New optimality principles for economic efficiency and equilibrium. *Journal of Optimization Theory and Applications* **75**, 221-264.
19. Farrell M J (1957). The measurement of productivity efficiency. *Journal of Royal Statistical Society A* **20**, 253-281.

(*Received October* 19, 2007 *Accepted July* 1, 2008)