# Distinguishing Weak and Strong Disposability among Undesirable Outputs in DEA: The Example of the Environmental Efficiency of Chinese Coal-Fired Power Plants

*Hongliang Yang and Michael Pollitt*

September 2007

CWPE 0741 and EPRG 0717

# Distinguishing Weak and Strong Disposability among Undesirable Outputs in DEA: The Example of the Environmental Efficiency of Chinese Coal-Fired Power Plants

Hongliang Yang[a*], Michael Pollitt[b]
[a,b] Judge Business School, University of Cambridge, Cambridge, CB2 1AG, UK,
h.yang.03@cantab.net, m.pollitt@jbs.cam.ac.uk

Abstract

Different from traditional efficiency research and previous studies considering undesirable outputs, this paper proposes models which distinguish weak and strong disposability features among various undesirable outputs based on the technical nature of the undesirable outputs. The paper illustrates the approach using a research sample covering 582 base-load Chinese coal-fired power plants in 2002. Our final results show that (1) imposing the technically correct disposability features on undesirable outputs makes a significant difference to the final efficiency evaluation. This suggests the necessity of properly distinguishing disposability features among undesirable outputs in efficiency models; (2) compared to their US and European counterparts, Chinese power plants relatively waste more resources. This suggests a great urgency for the Chinese electricity industry to improve its efficiency in coal-fired electricity generation sector.

*Subject classifications:* Economics: input-output analysis; Environment; Government: energy policies; Industries: electric; Statistics: nonparametric.
*Area of review:* Environment, Energy and Natural Resources
*JEL Codes*: D24, L94

## 1. Introduction

In spite of widespread recognition that researchers should credit DMUs for their provision of desirable or marketable outputs, and penalize them for their provision of undesirable outputs, no agreement has been reached with regards to how to incorporate undesirable outputs into a non-parametric efficiency model. In terms of disposability assumptions for undesirable outputs, some authors have assumed that undesirable outputs are weakly disposable (e.g. Fare et al., 1989, 1996; Tyteca, 1997), and some have assumed that undesirable outputs are strongly disposable (e.g.

* The authors are both members of the ESRC Electricity Policy Research Group whose financial support is acknowledged. Comments on an earlier draft by Stephen Littlechild and Emmanuel Thanassoulis are also acknowledged. The authors have benefited from comments received at presentations in Cambridge University. All remaining errors are their own.

Korhonen and Luptacik, 2004). When compared to traditional studies which ignore the issue of undesirable outputs, all of these efforts effectively broaden our understanding of the efficiencies of various production systems, in which desirable and undesirable outputs are jointly produced. However, because none of these papers distinguish undesirable outputs in terms of their specific technical features, the results of these papers might be, to some extent, misleading. The non-parametric DEA approach is selected in this paper as the main method of research. Previous literature on the inclusion of undesirable outputs is first summarized and discussed. Subsequently, six nonparametric models are constructed in the light of this discussion in order to measure the efficiency of Chinese coal-fired power plants using different ways of including undesirable outputs.

The Chinese electricity system is currently the second largest in the world, both in terms of installed generating capacity and generated electricity. Despite its growing importance and rapidly increasing scale, few quantitative analyses have been made of its efficiency. To the authors' knowledge, Lam and Shiu (2004) has so far been the only efficiency analysis study on the Chinese electricity generation sector using the DEA approach. However, this paper is more likely to be a snapshot than a complete analysis of the Chinese electricity industry. This is above all because the number of observations available is quite small. The data used comprises annual figures aggregated in terms of administrative provinces, and therefore there are only 30 annual observations available for analysis. Also, only traditional variables are included in the model, leaving the emissions of power plants unconsidered. Yet the large amount of coal-fired generating capacity in China has posed serious environmental problems. For example, in 2005 the total amount of $SO_2$ emissions from China ranked as number one in the world, and the total amount of $CO_2$ emissions ranked second only to US in absolute terms. In facing the increasing threat of environmental degradation, further environmentally aware studies are required when analyzing the efficiency of Chinese power plants.

The research sample used in this paper is a large 2002 sample which includes approximately 582 utility and non-utility coal-fired power plants, distributed across 31 provinces in the mainland of China. The aggregated installed capacity of the sample, mostly large grid-dispatched coal-fired power plants in China in 2002, is 211.71 GW. Besides traditional variables, the three main emissions of coal-fired electricity generation, namely $CO_2$, $SO_2$, and $NO_x$, are included in the analysis.

This paper is organized as follows. Section 2 contains a review of the current literature on how to include undesirable outputs in nonparametric models. Section 3 outlines the research methodology of this paper. Section 4 describes the research data. Section 5 reports the empirical results and section 6 concludes the paper.

# 2. Literature Review

## 2.1 Existing Literature

In reality the joint production of desirable and undesirable outputs, such as pollution, causes difficulties in the measurement of the overall performance of DMUs. Kopp et al. (1982) pointed out that an efficiency ranking derived from a frontier production function model may be misleading if it ignores differences in environmental restrictions across the research sample.

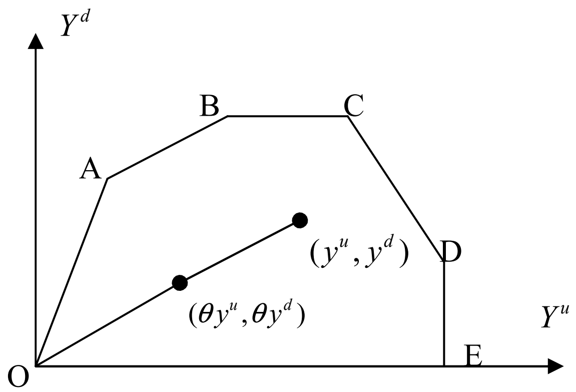
Pittman (1983) proposed that in holding the size of the overall capital stock constant, any increase in the capital devoted to the control of undesirable outputs will result in a corresponding reduction in the amount of desirable outputs. Therefore, any firm may potentially decrease both desirable and undesirable outputs synchronously without changing the inputs. He then extended Cave et al.'s (1982) trans-log multilateral superlative indexes to include undesirable outputs together with desirable outputs. In this model the desirable and undesirable outputs were treated asymmetrically. This new model was then used on a 1976 sample of 30 US mills which used pulp, together with capital, labour and energy, to produce paper and four undesirable pollutants. His results suggested that for industries in which undesirable outputs are important, the inclusion of undesirable outputs in the model results in some sizeable changes in the efficiency rankings of DMUs.

Fare et al. (1989) implemented the nonparametric approach to examine the same data set used by Pittman (1983). In this paper, Fare et al. distinguished between the strong and weak disposability of outputs[1]. Denote inputs by $x \in R_+$, desirable outputs by $y^d \in R_+$, and undesirable outputs by $y^u \in R_+$. To illustrate output disposability, the output set can be defined as

$$P(x) = \{(y^u, y^d) : (x, y^u, y^d) \in T\}.$$

Because $y^u$ cannot be reduced without cost, therefore, in Figure 1, under the weak disposability assumption of undesirable outputs, moving $(y^u, y^d)$ horizontally leftwards to the vertical axis on which $y^u = 0$ is not possible unless $y^d = 0$, and $(y^u, y^d)$ can only be proportionally scaled down to $(\theta y^u, \theta y^d)$ $(0 < \theta \le 1)$.

**Figure 1: Production Sets and Output Disposability**




Fare et al. (1989) then constructed several efficiency measures according to the different ways of including desirable and undesirable outputs asymmetrically. Their weak disposal reference technology is as follows:

$$P(x)_{Fare} = \{(y^u, y^d) : y^d \le Y^d \lambda, y^u = Y^u \lambda, x \ge X\lambda, \lambda \in R_+\}.$$

Fare et al.'s (1989) results are broadly similar to those of Pittman (1983). That is to say, the DMU's performance rankings are very sensitive to whether or not undesirable outputs are included. Therefore, traditional models might produce a misleading indication due to their exclusion of undesirable outputs.

Yaisawarng and Klein (1994) followed Fare et al.'s (1989) modelling strategy, but their paper was different from Fare et al. (1989) in two respects. Above all, they further distinguished desirable fixed inputs, e.g. capital. They claimed that this consideration avoided the labelling of a power plant as inefficient just because of excess capacity reserved for load variation. Secondly, they introduced into the model an undesirable input, namely sulphur content in the fuel. In so doing an input vector $x$ is redefined as $x = (x^v, x^f, x^u)$, in which $x^v$ refers to the variable desirable inputs, $x^f$ denotes the fixed desirable inputs, while $x^u$ represents the undesirable inputs. Yaisawarng and Klein's reference technology can be written as:

$$P(x^v, x^f, x^u)_{Y\&K} = \{(y^u, y^d) : y^d \le Y^d \lambda, y^u \gamma = Y^u \lambda, 1 \le \gamma < +\infty,$$
$$x^v \ge X^v \lambda, x^f \ge X^f \lambda, x^u \le X^u \lambda, \lambda \in R_+\}$$

where $\gamma$ is a scaling factor. In terms of Yaisawarng and Klein (1994), the scaling factor $\gamma$ is required on the constraint for undesirable outputs in order to ensure that weak disposability is satisfied in the case of VRS. This model was then used to examine the effect of $SO_2$ control on productivity change in US coal-fired power plants. By imposing strong disposability on sulphur content and weak disposability on $SO_2$, they acknowledged that a reduction of emissions must be accompanied by a reduction in desirable outputs, holding constant inputs.

Fare et al. (1996) introduced an environmental performance indicator by decomposing overall productivity into an environmental index and a productive efficiency index. Their disposability assumption and modelling methods were similar to those used in Fare et al. (1989). Models were used to examine two data sets of US fossil-fuel electric utilities. The results showed that when compared with that of the conventional model, the ranking of utilities obtained by the new model exhibits significant divergence.

Tyteca (1997) proposed three DEA models that include undesirable outputs. Tyteca defined a reference technology, as in Fare et al. (1989), in which undesirable outputs are assumed to be weakly disposable. The first model was designed to only reduce undesirable outputs in the technology set. It is called a undesirable output-oriented model and is equivalent to

$$H^o_{UO}(x, y^d, y^u) = \min\{\theta : (x, y^d, \theta y^u) \in T\}.$$

The above problem may be formulated as a group of linear programming functions to minimize the ratio of the weighted sum of undesirable outputs, to the weighted sum of desirable outputs less the weighted sum of inputs.

The second model was designed to minimise inputs along with undesirable outputs. Its functional form is

$$H_{IUO}^{o}(x, y^{d}, y^{u}) = \min\{\theta : (\theta x, y^{d}, \theta y^{u}) \in T\}$$

It is equivalent to a group of linear programming functions which minimise the ratio of the weighted sum of inputs and undesirable outputs over the desirable outputs. In the sense that both undesirable outputs and inputs should be decreased, this model in fact treated undesirable outputs as inputs.

Finally, the third model used a simpler form which no longer considered inputs. This model aimed to minimize the ratio of the weighted sum of undesirable outputs to that of desirable outputs. Its simpler functional form is

$$H_{NUO}^{o}(x, y^{d}, y^{u}) = \min\{\theta : (y^{d}, \theta y^{u}) \in T\}$$

These three models were later implemented on the same data set as that used in Fare et al. (1996). Given the similar models used in both papers, the results in Tyteca (1997) are also quite comparable with those achieved in Fare et al. (1996).

Korhonen and Luptacik (2004) borrowed the concept of eco-efficiency and used it to name the DEA efficiency indicator in the presence of undesirable outputs. They used two approaches to formulate their research models. In the first approach they divided the problem into two parts by measuring both technical efficiency, using a traditional DEA model, and so called ecological efficiency (as the weighted sum of desirable outputs to the weighted sum of undesirable outputs). Then both efficiency scores were taken as output variables for a new DEA model with an input equal to one.

In the second approach Korhonen and Luptacik treated undesirable outputs as inputs. Again, in the sense that given a certain number of desirable outputs, both inputs and undesirable outputs should be decreased. In terms of the different ways of including undesirable outputs when modelling, they constructed several DEA variants - similar to those used in Tyteca (1996, 1997). The first model was constructed to maximize the ratio of the weighted sum of desirable outputs less the weighted sum of undesirable outputs, to the weighted sum of inputs. The second model was structured to maximize the ratio of a weighted sum of desirable outputs, to the weighted sum of both inputs and undesirable outputs. The third model was built to maximize the ratio of the weighted sum of desirable outputs less the weighted sum of inputs, to the weighted sum of undesirable outputs. The above three models are input-oriented, while the fourth model is output-oriented and was also set up as the reciprocal of the second model. That is, the objective function was to minimize the ratio of the weighted sum of both inputs and undesirable outputs, to the weighted sum of desirable outputs. Although Korhonen and Luptacik (2004) did not explicitly state their disposability assumption, they in fact assumed that undesirable outputs are strongly disposable in their models. They implemented these models in order to examine the effects of an emissions reduction programme on 24 coal-fired power plants in a European country. In a different form from that used in other studies, the input variable used here was not physical unit of production factors, but the total cost of the emission reduction programme. Undesirable outputs included were dust, $NO_x$, and $SO_2$. Their results showed that all model variants lead to similar results, although the efficiency scores may differ.

## 2.2 Discussion

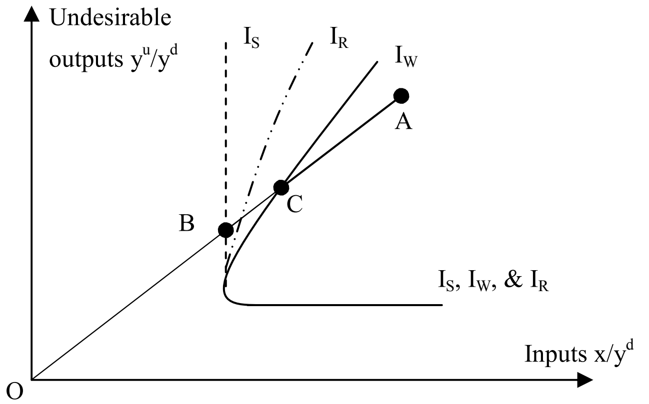
It can be seen from the above that when building efficiency measurement models, there has been no disagreement among authors on the selection of the strong disposability assumption for inputs and desirable outputs. Conflict occurs when considering the disposability assumption for undesirable outputs. Yet in spite of this conflict, overall they assumed a uniform disposability for all different undesirable outputs in a production system. Because none of these authors distinguish undesirable outputs in terms of their specific technical features, the nature of the uniform disposability assumption they made might therefore be arbitrary.

In reality the government or regulators may intervene in response to the joint production of undesirable outputs. For example, in the case of coal-fired electricity generation, power plants might be required to install desulphurization equipment in order to reduce $SO_2$ emissions. In terms of the specific characteristics of different pollutants and applicable abatement technologies, the disposal of undesirable outputs then becomes a two-sided problem. On one side, some of the undesirable outputs, such as $CO_2$ emissions, cannot be reduced using existing technology. Therefore, $CO_2$ emissions meet the definition of Fare and Grosskopf (2004) for 'Null-Joint Outputs' exactly[2]. Reducing $CO_2$ emissions will inevitably lead to a decrease in the amount of electricity generated. Equivalently, if electricity is generated, then some $CO_2$ emissions must also be produced. So for undesirable outputs like $CO_2$ emissions, weak disposability assumption is appropriate.

On the other side, some undesirable outputs can be disposed of. Although the disposal of these undesirable outputs may not be without cost, it is likely to be done with an acceptable increase in the cost of production. For example, it might first of all be likely for a coal-fired power plant to reduce $SO_2$ emissions by simply switching to low-sulphur coal. Alternatively, plants can employ wet flue gas desulphurization (FGD) systems which typically operate at more than 90% efficiency (Nolan, 1998; Kawatra and Eisele, 2001; Wang, 2002). Although the cleaning up of $SO_2$ emissions involves some input cost (or output reduction), a coal-fired power plant need not do this at the cost of reducing 90% of its electricity generation. In this situation, the Null-Joint relationship between emissions and desirable outputs is actually broken and the weak disposability assumption subsequently becomes inappropriate for undesirable outputs like $SO_2$ emissions. Therefore, we argue that a uniform disposability assumption may be arbitrary here because it conceals the specific technical features of different undesirable outputs.

These two sides to the disposal of undesirable outputs show that it is necessary for us to distinguish the technical features of various undesirable outputs before making any uniform and arbitrary disposability assumption. Figure 2 shows different frontiers under different disposability assumptions for undesirable outputs. $I_S$ and $I_W$ represent the frontiers under the strong and weak disposability assumptions respectively. The authors further hypothesize that, in reality, the co-incidence of strong and weak disposability among multiple undesirable outputs in a particular production process must lead to results somewhere in between those given by assuming all undesirable outputs have either strong or weak disposability. That is to say, in Figure 2 the real frontier ($I_R$) of a production system must be somewhere in between the strong and weak disposability extremes, and the real technical efficiency should be present in the interval between $OB/OA$ and $OC/OA$.

**Figure 2: Inputs – Undesirable Outputs Curves Under Different Disposability Assumptions**




## 3. Methodology

Assume we have $N$ (homogeneous) DMUs each using $M$ inputs to produce $P$ desirable outputs and $S$ undesirable outputs. Vectors $y_j^d$ and $y_j^u$ refer to the desirable and undesirable outputs of DMU $j$ respectively. Let $Y \in R_+$ be the output matrix, consisting of non-negative elements. Then the output matrix $Y$ can be decomposed as

$$Y = \begin{pmatrix} Y^d \\ Y^u \end{pmatrix},$$

where a $P \times N$ matrix $Y^d$ stands for desirable outputs and an $S \times N$ matrix $Y^u$ stands for undesirable outputs. Also, we define $x_j$ (an $M$-dimension vector) as the inputs consumed by DMU $j$, and $X \in R_+^{M \times N}$ as the input matrix. We assume that, given a certain amount of desirable output, we would like to use as little input as possible and to produce as little undesirable output as possible. Using $F_j(X, Y^d, Y^u)$ to represent the efficiency measurement for DMU $j$, then various DEA models can be set up as shown in the following subsections.

## 3.1 Model 1

In Model 1, following Fare et al. (1989), weak disposability is assumed for all undesirable outputs. The corresponding reference technology is then as follows:

$$T^{w} = \{(x, y^{d}, y^{u}) : y^{d} \leq Y^{d}\lambda, y^{u} = Y^{u}\lambda, x \geq X\lambda, \lambda \in R_{+}\}.$$

The efficiency measure of a DMU can be computed by solving the following linear programming function:

$$\begin{aligned} & F_{j}(X, Y^{d}, Y^{u}) = \min \theta \\ & s.t. \\ & y_{j}^{d} \leq Y^{d}\lambda \\ & \theta y_{j}^{u} = Y^{u}\lambda \\ & \theta x_{j} \geq X\lambda \\ & \lambda \in R_{+} \end{aligned}$$

## 3.2 Model 2

In Model 2, contrary to Model 1, strong disposability is assumed for all undesirable outputs. The strong disposal reference technology satisfying this assumption is therefore as follows:

$$T^{s} = \{(x, y^{d}, y^{u}) : y^{d} \leq Y^{d}\lambda, y^{u} \geq Y^{u}\lambda, x \geq X\lambda, \lambda \in R_{+}\}.$$

In line with Korhonen and Luptacik (2004), the corresponding efficiency measure is constructed to solve the following linear programming problem:

$$\begin{aligned} & F_{j}(X, Y^{d}, Y^{u}) = \min \theta \\ & s.t. \\ & y_{j}^{d} \leq Y^{d}\lambda \\ & \theta y_{j}^{u} \geq Y^{u}\lambda \\ & \theta x_{j} \geq X\lambda \\ & \lambda \in R_{+} \end{aligned}$$

## 3.3 Model 3

In Model 3, as discussed in the previous section, different disposability assumptions are used for different undesirable outputs in order to reflect the status quo of the pollution abatement technologies which are used in real life. In so doing, undesirable outputs with weak disposability are denoted by $y_w^u$, while undesirable outputs with strong disposability are denoted by $y_s^u$. The corresponding reference technology satisfying this assumption is therefore as follows:

$$T^{s\&w} = \{(x, y^d, y_w^u, y_s^u) : y^d \le Y^d\lambda, y_w^u = Y_w^u\lambda, y_s^u \ge Y_s^u\lambda, x \ge X\lambda, \lambda \in R_+\}.$$

Accordingly, the efficiency measure is as follows.

$$\begin{aligned}
&F_j(X, Y^d, Y_w^u, Y_s^u) = \min\theta \\
&s.t. \\
&y_j^d \le Y^d\lambda \\
&\theta y_{w,j}^u = Y_w^u\lambda \\
&\theta y_{s,j}^u \ge Y_s^u\lambda \\
&\theta x_j \ge X\lambda \\
&\lambda \in R_+
\end{aligned}$$

Up to this point Model 1, Model 2, and Model 3 have been constructed to examine the effects of different disposability assumptions on the efficiency measurement. From now on another three models will be set up to check the effects of desirable fixed inputs. There are mainly two considerations to be acknowledged for this setup. Firstly, following Yaisawarng and Klein (1994), this setup avoids labeling a power plant as inefficient because of load variation, something which can be found in many power plants. Secondly, the managers of power plants have, in reality, little control over generating capacity in the short run. Therefore, we can only seek radial reduction in the inputs which may be adjusted in the short term. This setup ensures a reduction in the quantity of labour and fuel needed, while holding a fixed capital quantity.

To facilitate our comparison Model 4, Model 5, and Model 6 use the same disposability assumptions as Model 1, Model 2, and Model 3, respectively. That is to say, the difference between Model 1 and Model 4 is whether or not some desirable fixed inputs are considered. The same is so for the other two model pairs, e.g. Model 2 and Model 5, and Model 3 and Model 6. We define $x_j^v$ as inputs whose values are varied in the report period, and $x_j^f$ as inputs whose values are fixed in the report period. The input matrix can then be interpreted as:

$$X = \begin{pmatrix} X^v \\ X^f \end{pmatrix},$$

where $X^v \in R_+$ and $X^f \in R_+$ refer to variable and fixed inputs respectively. Model 4, Model 5, and Model 6 are then formulated correspondingly.

## 3.4 Model 4

As is the same in Model 1, weak disposability is assumed for all undesirable outputs. The corresponding reference technology is then structured as

$$T^w = \{(x^v, x^f, y^d, y^u) : y^d \le Y^d \lambda, y^u = Y^u \lambda, x^v \ge X^v \lambda, \sum_{i=1}^{N} x^f \lambda_i \ge X^f \lambda, \lambda \in R_+\}.$$

The efficiency measurement of Model 4 can be computed by solving the following linear programming function[3],

$$
\begin{aligned}
& F_j(X^v, X^f, Y^d, Y^u) = \min \theta \\
& s.t. \\
& y_j^d \le Y^d \lambda \\
& \theta y_j^u = Y^u \lambda \\
& \theta x_j^v \ge X^v \lambda \\
& \sum_{i=1}^{N} x_j^f \lambda_i \ge X^f \lambda \\
& \lambda \in R_+
\end{aligned}
$$

where $\theta$ is a scalar equal to the efficiency score and $\lambda$ is a $N \times 1$ constant vector describing the weights of other DMUs used to construct the virtual frontier DMU.

## 3.5 Model 5

As in Model 2, strong disposability is assumed for all undesirable outputs. The strong disposal reference technology satisfying this assumption is as follows:

$$T^s = \{(x^v, x^f, y^d, y^u) : y^d \le Y^d \lambda, y^u \ge Y^u \lambda, x^v \ge X^v \lambda, \sum_{i=1}^{N} x^f \lambda_i \ge X^f \lambda, \lambda \in R_+\}.$$

The corresponding efficiency measure is then constructed to solve the following linear programming problem:

$$F_j(X^v, X^f, Y^d, Y^u) = \min\theta$$

$$s.t.$$

$$y_j^d \le Y^d\lambda$$

$$\theta y_j^u \ge Y^u\lambda$$

$$\theta x_j^v \ge X^v\lambda$$

$$\sum_{i=1}^{N} x_j^f \lambda_i \ge X^f\lambda$$

$$\lambda \in R_+$$

where $\theta$ is a scalar equal to the efficiency score and $\lambda$ is a $N\times 1$ constant vector describing the weights of other DMUs used to construct the virtual frontier DMU.

## 3.6 Model 6

Model 6 incorporates different disposability assumptions among undesirable outputs and considers the effects of desirable fixed inputs at the same time. The corresponding reference technology satisfying both requirements is therefore as follows:

$$T^{s\&w} = \{(x^v, x^f, y^d, y_w^u, y_s^u) : y^d \le Y^d\lambda, y_w^u = Y_w^u\lambda, y_s^u \ge Y_s^u\lambda, x^v \ge X^v\lambda, \sum_{i=1}^{N} x^f\lambda_i \ge X^f\lambda, \lambda \in R_+\}$$

Accordingly, the efficiency measure is constructed to solve the following linear programming problem:

$$F_j(X^v, X^f, Y^d, Y_w^u, Y_s^u) = \min\theta$$

$$s.t.$$

$$y_j^d \le Y^d\lambda$$

$$\theta y_{w,j}^u = Y_w^u\lambda$$

$$\theta y_{s,j}^u \ge Y_s^u\lambda$$

$$\theta x_j^v \ge X^v\lambda$$

$$\sum_{i=1}^{N} x_j^f \lambda_i \ge X^f\lambda$$

$$\lambda \in R_+$$

where $\theta$ is a scalar equal to the efficiency score and $\lambda$ is a $N\times 1$ constant vector describing the weights of other DMUs used to construct the virtual frontier DMU.

Based on the above six models, two different kinds of comparison can be made. Firstly, models 1, 2, and 3, and models 4, 5, and 6 form two respective comparison groups in terms of different disposability assumptions for undesirable outputs. Comparing the three sets of efficiency scores in a group demonstrates whether or not different disposability assumptions distort the values, and even the rankings, of the performance measurement. Secondly, models 1 and 4, models 2 and 5, and models 3 and 6, form three comparison pairs. Comparing each pair demonstrates the extent to which ignoring the effects of desirable fixed inputs varies the final efficiency results. Table 1 shows an overview of these models.

**Table 1: Overview of models**

| Variables | Traditional Model | Group 1 | | | Group 2 | | |
|---|---|---|---|---|---|---|---|
| | | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 |
| **Desirable output:** | | | | | | | |
| *Annual generation* | S | S | S | S | S | S | S |
| **Inputs:** | | | | | | | |
| *Installed capacity* | S | S | S | S | F | F | F |
| *Labour* | S | S | S | S | S | S | S |
| *Fuel* | S | S | S | S | S | S | S |
| **Undesirable output:** | | | | | | | |
| *$SO_2$ emissions* | -- | W | S | S | W | S | S |
| *$NO_x$ emissions* | -- | W | S | S | W | S | S |
| *$CO_2$ emissions* | -- | W | S | W | W | S | W |

Note: Inputs and desirable outputs are assumed to be strongly disposable by default. The signs 'W' and 'S' denote weak and strong disposability respectively. The sign 'F' denotes desirable fixed input and the sign '—' denotes exclusion of undesirable outputs in the traditional model.

# 4. Data and Variables

The research data sample used covers 582 Chinese coal-fired power plants during 2002, which mostly consisted of large grid-dispatched coal-fired power plants. The total installed capacity of the sample power plants is 211.71GW. The total annual generation is 1117.59 TWh. Data, such as installed capacity, annual fuel consumption (coal and oil), number of employees, annual electricity generation, heat rates, and quality of fuel, were collected by one of the author's through fieldwork in China during 2005 and 2006. Other data used, such as energy input and $CO_2$, $SO_2$, and $NO_x$ emissions, is calculated in terms of the *IPCC Reference Approach*, explained below.

## 4.1 Traditional Variables

Traditional variables used include electricity generated, capital, labour, and fuel. Electricity generated is considered a desirable output and is measured in the unit of MWh. Capital is measured by installed capacity *(*MW*)*. Labour is measured by number of employees. Quality of labour can be very different in terms of education, training, experience, etc. However, because it is hard to measure, in this research, we simply assume that there is no noticeable difference among workers. Because this number varies during the report period, the yearly average at the end of 2002 is adopted. Fuel is measured by energy (or heat) input. This is because in almost all Chinese coal-fired power plants oil-fired (sometimes gas-fired) equipment is also installed for boiler-preheating and standby purposes. The capacity of oil-fired or gas-fired equipment can be very different in terms of types of boiler and designs of combustion facilities. Commonly, given a certain loading on a boiler, the more oil or gas it uses, the lower the amount of coal consumed. Therefore, in order to make the final efficiency evaluation accurate and a comparison between plants meaningful, all coal, oil, and gas consumption are converted to energy (or heat) input which is measured in terajoules (TJ).

## 4.2 Undesirable Outputs

Undesirable variable refers to emissions from the electricity generation process. Coal is a combustible mineral composed primarily of carbon and hydrocarbon, along with other assorted elements including nitrogen and sulphur. Emissions from coal combustion mainly comprise $CO_2$, $SO_2$, $CH_4$, $N_2O$, $NO_x$, CO, and Non-methane volatile organic compounds (NMVOC). An accurate estimation of these emissions depends upon having knowledge of several interrelated factors, including combustion conditions, technology and emission control policies, as well as fuel characteristics. In general, the identification and quantification of emissions by fuel type is essential for the performance evaluation of power plants in this research. There are numerous different methods for estimating emissions. The methods used here are derived from the reference approaches of the *IPCC Guidelines for National Greenhouse Gas Inventories*. Emissions are estimated in terms of annual fuel consumption and average emission factors. On the basis of currently available data resources, the emissions considered are $CO_2$, $SO_2$, and $NO_x$. Table 2 presents the statistics of the variables used.

**Table 2: Descriptive Statistics of Research Sample**

| Variable | Unit | Mean | Maximum | Minimum | Std. Err. |
|---|---|---|---|---|---|
| **Desirable output:** | | | | | |
| *Annual Generation* | 1000 MWh | 1920 | 12423 | 17 | 2114 |
| **Inputs:** | | | | | |
| *Installed Capacity* | MW | 364 | 2520 | 12 | 400 |
| *Labour* | no. | 817 | 4627 | 107 | 680 |
| *Fuel* | TJ | 20715 | 124968 | 219 | 21194 |
| **Undesirable output:** | | | | | |
| *$SO_2$ Emissions* | tonne | 16493 | 194595 | 350 | 21054 |
| *$NO_x$ Emissions* | tonne | 6215 | 37490 | 66 | 6358 |
| *$CO_2$ Emissions* | 1000 tonnes | 1913 | 11541 | 20266 | 1957 |

Note: sample size = 582

# 5. Results

The application of this research includes the implementation of all models explained in section 3. For comparison purposes a ‘traditional model’, which only considers traditional variables (annual generation, installed capacity, labour and fuel) has also been built (Table 1). The calculation of all DEA models in this section is carried out by Matlab programs written by one of the authors. Table 3 shows the summary statistics of efficiency scores for the different models.

**Table 3: Summary Statistics of efficiency scores**

| Models | Traditional model | Group 1 | | | Group 2 | | |
|---|---|---|---|---|---|---|---|
| | | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 |
| Mean | 0.759 | 0.881 | 0.803 | 0.853 | 0.875 | 0.811 | 0.856 |
| Std. Err. | 0.115 | 0.108 | 0.119 | 0.115 | 0.110 | 0.121 | 0.114 |
| Min. | 0.391 | 0.548 | 0.404 | 0.527 | 0.425 | 0.381 | 0.405 |
| Max | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| No. of efficient DMUs | 21 | 114 | 40 | 69 | 118 | 57 | 78 |

## 5.1 Efficiency Scores

Table 3 shows that, above all, there is a substantial difference in efficiency scores between the traditional model and those including undesirable outputs. First of all, the traditional model tends to have a smaller efficiency score. The average efficiency score of the traditional model is 0.759 and the average efficiency scores of models 1 to 6 are located in the interval of 0.81-0.89. That is to say, in both situations a large inefficiency can be found in the sample power plants. Secondly, the traditional model tends to have less efficient DMUs. This agrees with the features of DEA – as variables included in DEA increase, the number of efficient DMUs will also increase. This result shows that the inclusion of undesirable outputs into DEA gives some advantages to those power plants which produce less emissions.

Secondly, Table 3 shows that in each comparison group the mean values of the models with weak and strong disposability assumptions respectively (Model 1 and Model 2 in Group 1, and Model 4 and Model 5 in Group 2) form two extremes on the efficiency score interval. Models with strong disposability assumptions for undesirable outputs (Model 2 in Group 1 and Model 5 in Group 2) achieve the lowest values in their efficiency scores. The efficiency scores of the models differentiating disposability features among undesirable outputs (Model 3 in Group 1 and Model 6 in Group 2) lie in the middle of the interval. This agrees with the discussion made in the section 2.

Thirdly, those models with a desirable fixed input arrangement (Group 2) tend to average lower efficiency scores than those without (Group 1). To some extent this supports Yaisawarng and Klein's (1994) introduction of fixed capital input. This also shows that the fixed input arrangement has a negative effect on the efficiency scores of DEA models (e.g. Banker and Morey, 1986). The reason for this is that the efficiency scores between Group 1 and Group 2 have a different interpretation. Group 1 scores indicate the extent to which all inputs can be proportionately reduced, whereas Group 2 scores indicate the extent to which only non-fixed inputs can be proportionately reduced.

Fourthly, models with a desirable fixed input arrangement (Group 2) tend to have more efficient DMUs than those without (Group 1) in terms of different disposability assumptions. Models with a weak disposability assumption for undesirable outputs have the largest number of efficient DMUs in both groups, and models with a strong disposability assumption have the least number of efficient DMUs.

## 5.2 Correlation Coefficients

Tables 4 and 5 exhibit the simple and the rank correlation coefficients of the efficiency scores respectively. Correlation coefficients and statistical tests were performed using the statistical package STATA.

**Table 4: Correlation of Efficiency Scores for Different Models**

| *Models* | *Traditional* | *Group 1* | | | *Group 2* | | |
|---|---|---|---|---|---|---|---|
| | | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | **Model 6** |
| **Traditional** | **1.0000** | | | | | | |
| **Model 1** | **0.5523** | **1.0000** | | | | | |
| **Model 2** | **0.9210** | **0.5707** | **1.0000** | | | | |
| **Model 3** | **0.6870** | **0.7635** | **0.7505** | **1.0000** | | | |
| **Model 4** | **0.4997** | **0.7723** | -- | -- | **1.0000** | | |
| **Model 5** | **0.7830** | -- | **0.8701** | -- | **0.6680** | **1.0000** | |
| **Model 6** | 0.5734 | -- | -- | 0.7640 | 0.7868 | 0.7600 | 1.0000 |

**Table 5: Spearman's Rank Correlation of Efficiency Scores**

| *Models* | *Traditional* | *Group 1* | | | *Group 2* | | |
|---|---|---|---|---|---|---|---|
| | | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | **Model 6** |
| **Traditional** | **1.0000** | | | | | | |
| **Model 1** | **0.4931** | **1.0000** | | | | | |
| **Model 2** | **0.9048** | **0.5247** | **1.0000** | | | | |
| **Model 3** | **0.6386** | **0.7476** | **0.7217** | **1.0000** | | | |
| **Model 4** | **0.4646** | **0.7897** | -- | -- | **1.0000** | | |
| **Model 5** | **0.7918** | -- | **0.8804** | -- | **0.6349** | **1.0000** | |
| **Model 6** | 0.5562 | -- | -- | 0.7725 | 0.7833 | 0.7442 | 1.0000 |

A high correlation coefficient between the two sets of efficiency scores generally indicates a high consistency for both sets. From Table 4 and Table 5 it can be seen that, firstly, the efficiency scores of Model 3 in Group 1 and Model 6 in Group 2 have higher correlation coefficients with other models which incorporate undesirable outputs. This corresponds with the fact that Model 3 and Model 6 incorporate both strong and weak disposability features for undesirable outputs.

Secondly, for each comparison pair the correlation coefficient is very high in both simple and rank correlations. This is to be expected given that the only difference between the two models in each pair is the inclusion of the fixed capital input. From another aspect, the high correlation coefficient also suggests that the introduction of desirable fixed inputs does not have as much influence on final efficiency scores as the choice of disposability assumption.

## 5.3 Rank-sum test

One question which can be asked here is whether or not the difference among various models is significantly important. This research uses the rank-sum test and significance test. This is because, firstly, the classic T-test assumes normality of the distributions, while the rank-sum test is distribution-free (Lehmann, 1975; Cooper et al., 2000). Since the theoretical distribution of efficiency scores in DEA is usually unknown, the use of the T-test in this context is not recommended. Secondly, the rank-sum test is a nonparametric technique by nature and this is consistent with the characteristics of DEA.

In line with the model specifications in the previous section, two kinds of rank-sum tests are performed in this research. The first kind tests the effects of different disposability arrangements in a group. As there are three efficiency score series in a group, the Kruskal-Wallis test is used to conduct the test (Lehmann, 1975; Brockett and Golany, 1996). The second kind of rank-sum test is used to test the effect of setting up fixed capital input. The Wilcoxon-Mann-Whitney Test has been selected to test the hypothesis of no difference in any two series of the efficiency scores of models. Table 6 and Table 7 show the test statistics of the Kruskal-Wallis test and Wilcoxon-Mann-Whitney Test respectively.

**Table 6: Kruskal-Wallis Test Statistic**

| ***Test statistic*** | ***Group 1*** | | | ***Group 2*** | | |
|---|---|---|---|---|---|---|
| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 |
| $\chi^2$ | 134.69*** | | | 92.708*** | | |
| $\chi^2$ **with ties** | 135.284*** | | | 93.200*** | | |

Note: *** denotes that the parameter is significantly important at 0.1%.

**Table 7: Wilcoxon-Mann-Whitney Test Statistic**

| *Models* | *Group 1* | | | *Group 2* | | |
|---|---|---|---|---|---|---|
| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | **Model 6** |
| **Model 1** | 0.000 | | | | | |
| **Model 2** | 11.386*** | 0.000 | | | | |
| **Model 3** | 4.409*** | -7.334*** | 0.000 | | | |
| **Model 4** | 0.890 | -- | -- | 0.000 | | |
| **Model 5** | -- | -0.979 | -- | 9.323*** | 0.000 | |
| **Model 6** | -- | -- | -0.442 | 3.068** | -6.589*** | 0.000 |

Note: ***, **, and * denote that the parameter is significantly important at 0.1%, 1%, and 5%, respectively.

From Table 6 it can be seen that the hypothesis of no difference within both groups can be rejected at 0.1% significance level. This indicates that different disposability arrangements for undesirable outputs do affect final efficiency measurements.

From Table 7 we can see that, firstly, in either Group 1 or 2 the hypothesis of no difference between any two models can be rejected at 0.1% significance level. This confirms the results of the Kruskal-Wallis test in Table 6. Secondly, test statistics suggest again that the introduction of desirable fixed inputs does not have as much influence on the final results of models. The test statistics for all three pairs (models 1 and 4, models 2 and 5, and models 3 and 6) cannot be rejected at any sensible level.

To summarize, the empirical results suggest that the assumption of a strong or weak disposability for undesirable outputs does influence the final efficiency measurement, while the effect of treating capital as a fixed input rather than a variable one is not very significant in the research sample of Chinese coal-fired power plants in operation during 2002.

# 6. Conclusion

Studies on how to incorporate undesirable outputs into performance measurements have been conducted for about two decades. However, no agreement has been reached regarding how to do this. Different authors have used different ways to formulate their efficiency measurement models (e.g. Fare et al., 1989, 1996; Yaisawarng and Klein, 1994; Tyteca, 1996, 1997; Korhonen and Luptacik, 2004). When compared to traditional studies these efforts effectively broaden our understanding of the efficiency of various production systems in which desirable and undesirable outputs are jointly produced. However, because none of these papers distinguish undesirable outputs in terms of their specific technical features, the authors suggest that assuming a uniform disposability assumption for various undesirable outputs in a production system might be arbitrary. Taking coal-fired power plants as an example, the authors point out that in reality some undesirable outputs might be strongly disposable and some might be weakly disposable.

In this paper, previous literature on the inclusion of undesirable outputs has been examined and summarized. The strengths of existing papers are combined in order to construct a set of new models. Based on the general guideline that given a certain amount of desirable output we would like to use as little input as possible and produce as little undesirable output as possible, six DEA models have been constructed to test two different effects: namely, the effects of different disposability arrangements for undesirable outputs and the effects of introducing fixed capital inputs into DEA models.

The empirical results show that, firstly, the strong or weak disposability assumption does affect the final efficiency measurement, while the influence of fixed capital input is not significant in this research sample. Secondly, the mean values of the efficiency scores under weak disposability are greater than those under strong disposability. Together they form two extremes to envelop reality. It is therefore necessary for us to distinguish disposability features on the basis of technical reality among various undesirable outputs before a more objective evaluation can be achieved.

This research not only contributes to the research methodology regarding how to incorporate undesirable outputs, but also entails various policy implications. This paper, first of all, attempts to find out the way to give a more objective efficiency evaluation for current coal-fired power plants in China by building and comparing seven different research models. Results show that whether considering undesirable outputs and whether distinguishing disposability features among undesirable outputs pose a significant difference in the final efficiency evaluation. Then, which model would be the best for adoption for assessing the environmental efficiency of Chinese coal-fired power plants? The authors support models distinguishing disposability features among undesirable outputs. Against the increasing environment concern, it makes sense for us to give some priority to those coal-fired power plants which are more effective in emissions control. The research models distinguishing disposability features can be used as a starting point for this purpose.

Also, compared to similar studies for other countries, this paper shows that the efficiency divergence in Chinese coal-fired power plants is much bigger. For example, after considering the effects of $SO_2$ emissions, Yaisawarng and Klein (1994) reported that on average the inefficiency was found to be less than 8% in their research sample, which covered 61 US coal-fired power plants. When including DUST, $NO_x$, and $SO_2$ as undesirable outputs in the efficiency models, Korhonen and Luptacik (2004) obtained an average inefficiency of around 7% using a research sample with 24 coal-fired power plants in a European country. However, in this paper the average inefficiency is found to be about 12% in models assuming weak disposability for undesirable outputs, about 20% in models assuming strong disposability for undesirable outputs, and about 15% in models distinguishing disposability features among undesirable outputs. This indicates that, relatively, Chinese coal-fired power plants waste more resources than their counterparts in US and Europe. The existence of the large inefficiency in current facilities suggests a great urgency for the Chinese electricity industry to improve its efficiency in coal-fired electricity generation sector. Statistics shows that in 2003 electricity demand increase in China was about 15% (SPIN, 2004). This implies that if Chinese coal-fired power plants can operate more efficiently, there is no need for China to install that much coal-fired generating capacity.

Furthermore, this paper also tests the effect of fixed capital input. Results show that this effect is not very significant. However, given that almost all coal-fired power plants were relatively fully utilized in an environment of electricity shortage in 2002, the result of insignificant effect of fixed capital input should be used very cautiously and needs further examination in the future.

## Endnotes

[1] Strong disposability of outputs implies that given an input vector $x$, if an output vector $y$ can be produced, then $y^*$ can also be produced as long as $y^* \le y$. Strong disposability is also called free disposability. Weak disposability of outputs means that if $y$ can be produced, then $\theta y (0 \le \theta \le 1)$ can also be produced proportionally.

[2] In terms of Fare and Grosskopf (2004), we can say that the desirable output vector $y^d$ is **Null-Joint** with the undesirable outputs $y^u$ if $(y^u, y^d) \in$ output set $P(x)$, and $y^u = 0$ then $y^d = 0$. That is, if $(y^u, y^d)$ is feasible and there are no undesirable outputs produced, then under null jointness no desirable outputs can be produced.

[3] There are two factors worthy of emphasizing here. Above all, the constraint for the fixed capital input is set in terms of constraint (32) of Banker and Morey (1986), except for some notation differences. In Banker and Morey's, the DMU under evaluation was represented by '0', but in this paper, it is represented by 'j'; in Banker and Morey's, the summation of $\lambda$ was over 'j', but in this paper, it is over 'i'. Also, the constraint for the fixed capital input is set under CRS in this paper. If VRS is assumed, then $\sum_{i=1}^{N} \lambda_i = 1$, and then this constraint will collapse to the following format:

$$x_j^f \ge X^f \lambda .$$

That is to say, if VRS is assumed, Model 4 will take the form of the efficiency measure from (19)-(24) in Banker and Morey (1986).

# References


Brockett, P. and Golany, B. 1996. Using Rank Statistics for Determining Programmatic Efficiency Differences in Data Envelopment Analysis. *Management Science.* 42(3). 466-472.

Caves, D. Christensen, L. and Diewert, W. 1982. Multilateral Comparisons of Output, Input, and Productivity Using Superlative Index Numbers. *Economic Journal.* 92. 73-86.

Charnes, A., Cooper, W., and Rhodes, E. 1978. Measuring Efficiency of Decision Making Units. *European Journal of Operational Research.* 2. 429-444.

Cooper, W., Seiford, L., and Tone, K. 2000. *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solvers Software*. London: Kluwer.

Fare, R., and Grosskopf, S. 2004. *New Directions: Efficiency and Productivity.* Kluwer Academic Publishers, Boston.

Fare, R., Grosskopf, S., Lovell, K., and Pasurka, C. 1989. Multilateral Productivity Comparisons When Some Outputs Are Undesirable: A Nonparametric Approach. *The Review of Economics and Statistics.* 71(1).

Fare, F., Grosskopf, S., and Tyteca, D. 1996. An Activity Analysis Model of the Environmental Performance of Firms – Application To Fossil-fuel-fired Electric Utilities. *Econlogical Economics.* 18. 161-175.

Fare, R., Grosskopf, S. and Roos, P. 1997. Malmquist Productivity Indexes: A Survey of Theory and Practice. In Fare, R., Grosskopf, S. and Russell, R. (ed.), *Index Numbers: Essays in Honour of Sten Malmquist.* Kluwer Academic Publishers, Boston.

Grosskopf, S. 1993. Efficiency and Productivity. In Fried, H. Lovell, C. and Schmidt, S. (ed.). *The Measurement of Productive Efficiency: Techniques and Applications*. Oxford University Press, New York, 160-194.

Kopp, R., Smith, V., and Vaughan, W. 1982. Stochastic Cost Frontiers and Perceived Technical Inefficiency. In Smith, K. (eds) *Advances in Applied Micro-Economics.* 2. JAI Press.

Korhonen, P. and Luptacik, M. 2004. Eco-efficiency Analysis of Power Plants: An Extension of Data Envelopment Analysis. *European Journal of Operational Research.* 154. 437- 446.

Lam, P. and Shiu, A. 2004. Efficiency and Productivity of China's Thermal Power Generation. *Review of Industrial Organization.* 24: 73-93.

Nolan, Paul. 1998. Flue Gas Desulphurization Technology for China. *International Electric for China.* Babcock & Wilcox, Ohio, USA.

Pittman, R. 1983. Multilateral Productivity Comparisons with Undesirable Outputs. *The Economic Journal.* 93. 883-891.

State Environmental Protection Administration of China (SEPAC). 2006. *China Forum of Environmental Journalists.* Access on 8 August 8, 2006 at http://www.cfej.net/news.

Tyteca, D. 1996. On the Measurement of the Environmental Performance of Firms – A Literature Review and A Productive Efficiency Perspective. *Journal of Environmental Management*. 46. 281-308.

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.

Wang, Zhixuan 2002. *Environmental Consequence and Control Method of Sulphur Dioxide of Coal-fired Power Plants* (Huo Dian Er Yang Hua Liu Huan Jing Ying Xiang Yu Kong Zhi Dui Ce (In Chinese Pinyin)). China Environmental Science Press, Beijing.

Yaisawarng, S. and Klein, J. 1994. The Effects of Sulfur Dioxide Controls on Productivity Change in the US Electric Power Industry. *The Review of Economics and Statistics.* 76(3). 447-460.