European Journal of Operational Research 197 (2009) 1095–1105

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Incorporating both undesirable outputs and uncontrollable variables into DEA: The performance of Chinese coal-fired power plants ☆

Hongliang Yang *, Michael Pollitt

Judge Business School, University of Cambridge, Cambridge, CB2 1AG, UK

ARTICLE INFO

Article history:
Received 16 January 2007
Accepted 5 December 2007
Available online 13 March 2008

Keywords:
Data envelopment analysis
Performance measurement
Technical efficiency
Electricity

ABSTRACT

There are two difficulties in doing an objective evaluation of the performance of decision making units (DMUs). The first one is how to treat undesirable outputs jointly produced with the desirable outputs, and the second one is how to treat uncontrollable variables, which often capture the impact of the operating environment. Given difficulties in both model construction and data availability, very few published papers simultaneously consider the above two problems. This article attempts to do so by proposing six DEA-based performance evaluation models based on a research sample of the Chinese coal-fired power plants. The finding of this paper not only contributes for the performance measurement methodology, but also has policy implications for the Chinese coal-fired power sector.

1. Introduction

There are two difficulties in doing an objective evaluation of the performance of decision making units (DMUs). The first difficulty is how to treat undesirable outputs jointly produced with desirable outputs. Traditional literature only values the desirable and simply ignores the undesirable. However, ignorance of the undesirable is equal to saying that they have no value in the final evaluation and may present misleading results. It is therefore necessary to credit DMUs for their provision of desirables and penalize them for their provision of undesirables. The second difficulty is how to treat uncontrollable variables, which often reflect the impact of the operating environment. Generally, the management can decide on some controllable factors internal to production activities, while the impact of the operating environment is out of the control of the management. Traditional studies which have constructed research models using controllable factors only, implicitly assume that all the inefficiencies of DMUs are caused by bad management. Since the impact of uncontrollable variables is not filtered out, the evaluation of those DMUs in an adverse operating environment will be underestimated. However, given difficulties in both model construction and data availability, there are very few published papers which consider both problems simultaneously.

This article attempts to solve the above issue by setting up six DEA-based performance evaluation models for a production process which produces both desirable and undesirable outputs. However, unlike parametric methods, the inclusion of undesirable outputs and uncontrollable variables in DEA is still a nascent field of research. Previous works incorporating both of the variables are summarized. The models used include the basic model, the one-stage model, the two-stage model, the three-stage model and the four-stage model, following the characterisation offered in Pastor (2002). To the authors' knowledge, no other empirical studies on performance measurement or efficiency analysis have used all of these models in one paper.

The Chinese electricity sector is now the second largest electricity system in the world, both in terms of installed capacity and generation. In 2005 its total installed generation capacity was about 508.4 GW, of which about 384.1 GW was thermal generating capacity with more than 90% of that thermal capacity being coal-fired (Zhang, 2006). The wide usage of coal-fired capacity has posed many serious environmental problems and become a world concern. 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. However, despite the growing size of the Chinese coal-fired electricity system and increasing environmental concerns, few empirical quantitative analyses of the efficiency of Chinese coal-fired power plants have been completed. In this paper, a research sample of Chinese coal-fired power plants during 2002 is used. Data in the sample covers both traditional inputs and outputs and undesirable outputs (e.g. SO_2 emissions) and some carefully selected uncontrollable variables. This research sample allows

☆ The authors are both members of the ESRC Electricity Policy Research Group whose financial support is acknowledged. They have benefited from comments received at presentations in Cambridge University. All remaining errors are their own.

* Corresponding author. Tel.: +44 (0) 1223 766638; fax: +44 (0) 1223 339701.
E-mail addresses: h.yang.03@cantab.net (H. Yang), m.pollitt@jbs.cam.ac.uk (M. Pollitt).

0377-2217/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejor.2007.12.052

the authors to examine the impact of uncontrollable variables on the performance of coal-fired power plants together with undesirable outputs, and also to test the validity of different models constructed in the paper.

This paper is organized as follows: Section 2 reviews existing literature on the inclusion of undesirable outputs and uncontrollable variables; Section 3 explains the research methodology (the different models proposed for incorporating undesirable outputs and uncontrollable variables are laid out here); Section 4 describes the research data; Section 5 summarizes the results and Section 6 concludes the paper.

2. Literature review

2.1. Inclusion of undesirable outputs

Since Charnes et al. (1978) DEA has been widely used to measure the performance of various kinds of DMUs. In the case of power plants, examples from different eras include Färe et al. (1985), Pollitt (1995), Coelli (1997) and Olatubi and Dismukes (2000). Within these studies traditional inputs are used to produce desirable or marketable outputs. Such ignoring of undesirable outputs might produce misleading results. However, Fare et al. (1989) implemented the nonparametric approach on a 1976 data set of 30 US mills which use pulp and three other inputs in order to produce paper and four pollutants. In their research they assumed weak disposability[1] for undesirable outputs. Their results showed that the performance rankings of DMUs turned out to be very sensitive to whether or not undesirable outputs were included. That is to say, traditional DEA models might give us a biased indication of where we stand. Other studies exhibit similar results (e.g. Pittman, 1983; Tyteca, 1996, 1997). Thus, given the fact that undesirable outputs are jointly produced with desirable outputs, it makes sense for us to credit a DMU for its provision of desirable outputs and to penalize it for its production of emissions when evaluating its performance.

Yaisawarng and Klein (1994) constructed a DEA model to measure the effects of SO_2 control on the efficiency change of US coal-fired power plants in the 1980s. Following Fare et al. (1989), they assumed weak disposability for undesirable outputs and further extended their DEA model to include an undesirable input, namely sulphur content in the fuel. However, because SO_2 emissions can only come from the combustion of sulphur and as this is expected to be highly correlative with sulphur content in the fuel, the inclusion of both variables synchronously in the DEA is not necessary. Due to the nature of DEA modeling, for any fixed sample size, increasing the number of variables results in higher efficiency scores and more efficient units. Therefore, the inclusion of both variables may actually reduce the discriminating power of the DEA model.

Fare et al. (1996) introduced an environmental performance indicator by decomposing overall productivity into an environmental index and a productive efficiency index. The authors adopted disposability assumption and DEA modelling techniques similar to those in Fare et al. (1989). Models were then used to examine two data sets of US fossil-fired electric utilities. The results showed that, when compared with those of the traditional model, the ranking of utilities obtained by the new model was significantly different.

Tyteca (1996) provided a comprehensive survey of the previous literature on environmental performance measurement. He proposed three DEA variations in terms of how undesirable outputs are included. These three models were later implemented in Tyteca (1997) on the same data set as that used in Fare et al. (1996). Given the similar models constructed in both papers, the results of both papers are also quite comparable.

Korhonen and Luptacik (2004) used several variants of DEA models to measure the eco-efficiency of 24 coal-fired power plants in a European country. Their modelling methods were quite similar to those used in Tyteca (1996, 1997). They treated emissions directly as inputs in the sense that given a certain amount of desirable output, both inputs and undesirable outputs should be decreased. The results showed that all model variants lead to similar results. When compared to previous studies this paper provided a deeper insight on the efficiency of power plants.

In total, all the above explorations have effectively broadened our understanding of efficiency evaluation of DMUs. Based on the above, the basic DEA model of this paper is constructed. Following Korhonen and Luptacik (2004), undesirable outputs are included like inputs.

2.2. Incorporating uncontrollable variables into DEA

Uncontrollable variables here refer to those factors which influence the performance of DMUs and are at the same time out of the control of the management. We notice that many authors prefer to use the term environmental variable to express the same meaning (e.g. Coelli et al., 2005; Fried et al., 1999, 2002; Pastor, 2002). However, in order to avoid any unnecessary confusion with environmental emissions, this paper uses the uncontrollable variable term instead. Previous works which incorporate uncontrollable variables into DEA can be broadly classified as follows: separation models, one-stage models, two-stage models, three-stage models and four-stage models. In each case the basic aim is to level the playing ground for all players. In the rest of this section we sketch the different models and in Section 3 we characterize them formally.

2.2.1. Separation model

The separation model stratifies the research sample in terms of categorical variables which capture the characteristics of the operating environment (e.g. public vs. private ownership). DEA frontiers are then constructed for each category, respectively. Despite being easy to interpret and apply, it requires an advance selection of the most influencing features of the operating environment and can only be used for categorical variables. Also, because it greatly reduces the comparison set its implementation lessens the discriminating power of DEA. Some inefficient DMUs may become efficient after the research sample is subdivided. In addition, because DMU efficiency scores are calculated against the respective DEA frontiers, the comparison of efficiency scores across sub-samples becomes meaningless. Studies using this model include Charnes et al. (1981), Grosskopf and Valdmanis (1987), Banker et al. (1990) and Fizel and Nunnikhoven (1992).

2.2.2. One-stage model

The one-stage model directly includes uncontrollable variables in its linear functions, along with traditional inputs and outputs. This model can make use of the capability of DEA to accommodate multiple variables. However, it also has some shortcomings. Firstly, in line with traditional DEA models, it assumes that all uncontrollable variables can be radially altered. This assumption might not be reasonable for some features of the operating environment. A variation is proposed in order to solve this problem, in which uncontrollable variables are either held at a constant or cannot be reduced if cost increases or cannot be increased if cost decreases in the calculation. Although this variation precludes the original

[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 $y^* \leqslant y$. Strong disposability is also called free disposability. Weak disposability of outputs means that if y can be produced, and then $\theta y (0 \leqslant \theta \leqslant 1)$ can also be produced.

arbitrary assumption, the one-stage model still requires a prior decision regarding the influence direction of an uncontrollable variable. Secondly, as the number of variables included increases, the number of efficient DMUs is expected to increase as well. Examples using this approach include Banker and Morey (1986) and McCarty and Yaisawarng (1993).

2.2.3. Two-stage model

The two-stage model starts with a standard DEA model based on traditional inputs and outputs in the first stage, and regresses the efficiency scores of the first stage against a set of selected uncontrollable variables in the second stage. Because the efficiency scores of the first stage are confined in the interval (0, 1], the use of limited dependent variable regression techniques, such as Tobit or an exponential function model, is preferable. This model is easy to use and easy to interpret, and is also capable of accommodating both continuous and categorical uncontrollable variables without increasing the number of efficient DMUs. Furthermore, it does not require any prior knowledge regarding the influence direction of an uncontrollable variable. However, this model has some disadvantages as well. Firstly, it ignores the information contained in the input slacks or output surpluses. Secondly, if the variables used in the first stage are highly correlated with the variables used in the second stage, then the results may bias the parameter estimates regarding the impact of uncontrollable variables on efficiency (Coelli et al., 2005; Simar and Wilson, 2007). Examples of research using this approach are McCarty and Yaisawarng (1993), Fried et al. (1993), and Pollitt (1995).

2.2.4. Three-stage model

The three-stage model was first applied by Fried et al. (2002) on a 1993 sample of US hospital-affiliated nursing homes. The first stage comprises a standard DEA using traditional inputs and outputs. In the second stage stochastic frontier analysis (SFA) is used to regress the input slacks (radial plus non-radial) of the first stage against a set of selected uncontrollable variables. In this stage the total input slacks (radial plus non-radial) are decomposed into three parts: a part attributable to uncontrollable impacts, a part attributable to management inefficiency, and a part attributable to statistical noise. The stochastic slack frontier constructed here can be interpreted as the minimum slacks which can be achieved in a noisy environment. Based on the estimated coefficients, the inputs are then adjusted accordingly. The aim of this stage is to obtain slacks filtered for the impact of uncontrollable variables. In the third stage, DEA is repeated using the adjusted input values. The merits of this model are straightforward. First of all, it can thoroughly decompose input slacks and make best use of the information contained in the input slacks. Secondly, it can accommodate multiple uncontrollable variables and does not require any prior understanding with regards to their influence direction on the efficiency scores of the DMUs. Yet the cost of this model is high in terms of time and computation requirements.

2.2.5. Four-stage model

The four-stage model was introduced by Fried et al. (1999) to measure the impact of uncontrollable variables on DMU efficiency. In the first stage, a standard DEA is constructed using traditional inputs and outputs. In the second stage, total input slacks (radial plus non-radial) is regressed using Tobit against selected uncontrollable variables. In the third stage, parameters estimated in the second stage are used to estimate allowable input slacks. Then the values of primary inputs are adjusted accordingly. In the fourth stage, the DEA is repeated using the adjusted input values.

The four-stage model shares its modelling philosophy with the three-stage model. That is, DMUs operating in relatively unfavourable environments are disadvantaged in the traditional DEA model. Therefore, levelling of the playing field is necessary for an objective performance evaluation. The major difference between these two models is the functional form used in the regression of the second-stage. SFA is selected in the third-stage model, while the Tobit regression is used in the four-stage model. The four-stage model shares similar advantages and disadvantages with the three-stage model. However, when compared with the three-stage model, the four-stage model can only adjust inputs to account for environmental impact and not for statistical noise.

It is clear from the above that each model has its own advantages and disadvantages (Table 1). As such it would be right to say that the three-stage model is the most sophisticated model in terms of methodology. With the exception of the separation model, which is discarded due to its shortcomings, all of the other models will be used and compared in this paper.

3. Methodology

3.1. Basic DEA model

Assume that we have N (homogeneous) DMUs each producing P desirable outputs and $\boldsymbol{S}$ undesirable outputs while using M inputs. Let $Y \in R_+$ be the output matrix consisting of non-negative elements, and vectors y_j^d and y_j^u refer to the desirable and undesirable outputs of DMU j, respectively. Then the output matrix Y can be decomposed into two parts:

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

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, define $X \in R_+^{M\times N}$ as the input matrix and $F_j(X, Y^d, Y^u)$ as the input-oriented efficiency measurement for DMU j. Assume we like to produce as much as possible of a desirable output and as little as possible of an undesirable output. Based on the assumption of constant return to scale (CRS), an input-oriented DEA model can be formulated in the following format[2]:

$$\begin{aligned} F_j(X, Y^d, Y^u) = \quad & \text{Min} \quad \theta \\ & \text{s.t.} \quad Y^d\lambda \geqslant y_j^d, \\ & \qquad Y^u\lambda \leqslant \theta y_j^u, \\ & \qquad X\lambda \leqslant \theta x_j, \\ & \qquad \lambda \geqslant 0, \quad j = 1, \ldots, N, \end{aligned} \tag{1}$$

where λ is an $(N \times 1)$ vector of coefficients which represents the intensity levels for DMUs in the construction of the reference efficiency frontier.

Note that the convexity constraint $N1'\lambda = 1$ is not included in the above model. There are two reasons for this. Firstly, under a VRS frontier a DMU is only benchmarked against DMUs of a similar size (Coelli et al., 2005), it is therefore arguable whether or not a DMU's efficiency score under a VRS frontier effectively reflects its performance relative to the best practice in the industry. Secondly, the constraint might screen some of the effects of uncontrollable variables. In particular, the constraint is very likely to undermine the use of scale variables in multistage regression analysis. Such analysis enables the statistical testing of scale effects. The authors will introduce scale variables in this way in the latter part of the paper.

[2] The disposability of undesirable outputs has two facets. In coal-fired electricity generation, on the one hand, some undesirable outputs, e.g. CO_2 emissions, can only be weakly disposable using the existing technology; on the other hand, some undesirable outputs, e.g. SO_2 emissions, can be strongly disposable (Yang and Pollitt, 2007). Because in this paper only SO_2 emissions are considered, for simplicity the authors only include a constraint for undesirable outputs with strong disposability.

Table 1
Inclusion of uncontrollable variables

Research approach	Methods	Advantages and disadvantages	Example literature
Separation approach	Subdivide the research sample into sub-samples according to different uncontrollable variables	*Advantages:* Easy to interpret and apply *Disadvantages:* (1) Can only be used for a categorical variable each time (2) Lessens the discriminating power of DEA	Charnes et al. (1981),Banker and Morey (1986),Grosskopf and Valdmanis (1987), Banker et al. (1990), Fizel and Nunnikhoven (1992)
One-stage model	Include uncontrollable variables in DEA together with traditional inputs and outputs	*Advantages:* (1) Easy to interpret and apply (2) Able to accommodate multiple variables *Disadvantages:* (1) Requires a prior understanding of the influence direction of an uncontrollable variable (2) Some inefficient DMUs might become efficient as the number of the uncontrollable variables included increases	Banker and Morey (1986), McCarty and Yaisawarng (1993),
Two-stage model	(1) Construct a traditional DEA in the first stage (2) Regress efficiency scores from the first stage against a set of uncontrollable variables in the second stage	*Advantages:* (1) Easy to apply and interpret (2) Able to accommodate continuous and categorical variables without increasing the number of efficient DMUs (3) Does not require a prior understanding of the influence direction of each uncontrollable variable *Disadvantages:* (1) May bias the parameter estimates regarding the impact of uncontrollable variables on efficiency (2) If using OLS in the second stage, the corrected efficiency scores might not be between 0 and 1	McCarty and Yaisawarng (1993), Fried et al. (1993), Pollitt (1995)
Three-stage model	(1) Construct a traditional DEA in the first stage (2) Use SFA to estimate the impact of uncontrollable variables and statistical noise, and adjust the input values accordingly (3) Repeat the DEA in the third stage using the adjusted input values	*Advantages:* (1) Easy to understand (2) Able to accommodate many variables without requiring a prior understanding of their influence direction (3) Able to capture the information contained in the input slack *Disadvantages:* High cost of time and calculation	Fried et al. (2002)
Four-stage model	(1) Construct a traditional DEA in the first stage (2) Use Tobit to estimate the impact of uncontrollable variables in the second stage (3) Adjust the input values in the third stage (4) Repeat the DEA in the last stage using the adjusted input values	Shares similar advantages and disadvantages with the three-stage model. Compared to the three-stage model, it can only adjust the input values to account for uncontrollable variables and not for statistical noise	Fried et al. (1999)

So in a different approach to that used in Fried et al. (1999, 2002), both of which used VRS DEA as their first-stage technology, this research adopts CRS DEA instead.

3.2. Inclusion of uncontrollable variables

Assume that Z_j is a vector of uncontrollable variables characterizing the operating environment for unit j. Based on the explanation in the previous section, different models are formulated to include Z_j.

3.2.1. One-stage model

The one-stage model requires a prior differentiation on the influence of Z_j. Assume that Z_j^+ and Z_j^- are the vectors of uncontrollable variables for unit j with a positive and negative impact on performance respectively. The one-stage model can then be constructed as

$$F_j(X, Y^d, Y^u) = \begin{array}{ll} \text{Min} & \theta \\ \text{s.t.} & Y^d\lambda \geqslant y_j^d, \\ & Y^u\lambda \leqslant \theta y_j^u, \\ & X\lambda \leqslant \theta x_j, \\ & Z^+\lambda \leqslant Z_j^+, \\ & Z^-\lambda \geqslant Z_j^-, \\ & \lambda \geqslant 0, \quad j = 1, \ldots, N. \end{array} \tag{2}$$

3.2.2. Two-stage model

In the second stage of the two-stage model, efficiency scores from the basic model are regressed on a set of uncontrollable variables. In order to constrain the corrected efficiency scores within an interval between 0 and 1, this study follows Pastor (2002) in its use of two different regression techniques, one being Tobit regression and the other being non-linear logistic regression. In the latter case, the non-linear function is structured as follows:

$$\theta_j = \frac{\exp(Z_j; \beta)}{1 + \exp(Z_j; \beta)} + \varepsilon_j. \tag{3}$$

The parameters of both the Tobit regression and the logistic regression allow the efficiency scores of individual units to be adjusted to common levels of uncontrollable variables in order to facilitate proper comparison. The authors do this by taking the sum of the estimated impact of all uncontrollable variables for each DMU and adjusting each DMU's efficiency score to the average level of the estimated impact of the uncontrollable variables. For example, DMUs with unfavorable operating conditions have their efficiency scores adjusted upwards to reflect average operating environment.

Table 2
Descriptive statistics of power plants with uncontrollable variables

Variable	Unit	Mean	Maximum	Minimum	Standard error
Desirable output:					
Annual generation	1000 MWh	2614.26	12422.77	59.85	2262.41
Inputs:					
Installed capacity	MW	482.69	2400	12	407.27
Labour	No.	801	3674	136	645
Fuel	TJ	27073.73	124968	1121	22270.44
Undesirable output:					
SO_2 emissions	Tonnes	24110.14	194595	461	27538.53
Uncontrollable variables:					
Vintage	Year	10.19	43	1	8.09
Calorific value of coal	GJ/tonne	22.86	28.68	12.49	2.43
Scale 1 (91 plants)	0 < Scale < 200 MW	–	1	0	–
Scale 2 (115 plants)	200 MW ⩽ Scale < 400 MW	–	1	0	–
Scale 3 (15 plants)	400 MW ⩽ Scale	–	1	0	–
CHP (46 plants)	1 = Yes; 0 = No	–	1	0	–

Note: sample size = 221.

3.2.3. Three-stage model

The first stage of the three-stage model is conducted using the basic model, which provides the initial performance evaluation for each unit.

In the second stage SFA is used to decompose stage-one slacks. Since slacks cannot be aggregated across non-commensurate variables, SFA slack regressions[3] are conducted for different inputs and undesirable outputs, which are all treated like inputs in the model. The M separate SFA regressions take the following general form. For the mth input

$$S_{mj} = f^m(Z_j;\beta^m) + v_{mj} + u_{mj},\quad m = 1,2,\&,M;\quad j = 1,2,\&,N, \tag{4}$$

where $S_{mj} = x_{mj} - X_m\lambda$ are the stage-one slacks in the usage of the mth input for the jth DMU, including radial and non-radial slacks. Z_j is the vector of uncontrollable variables for the jth DMU. β^m is a vector of coefficient. We also assume that the $v_{mj} \sim N(0,\sigma^2_{vm})$ reflects statistical noise and the $u_{mj} \sim N^+(\mu^m,\sigma^2_{um})$ reflects managerial inefficiency. Estimates of coefficients in the function (4) make it possible to measure the contributions of different factors on the input slacks. $\hat{\beta}^m$ indicates the contribution of each uncontrollable variable, while $\hat{v}_{mj}$ and $\hat{u}_{mj}$ explain the effects of statistical noise and managerial inefficiency. Define $\gamma^m = \sigma^2_{um}/(\sigma^2_{vm} + \sigma^2_{um})$. While $\gamma^m \to 1$, the impact of managerial inefficiency dominates that of statistical noise, and while $\gamma^m \to 0$, statistical noise plays the dominant role.

In terms of the results of the SFA analysis, inputs are then adjusted to reflect the impact of uncontrollable variables and statistical noise. The inputs of those DMUs which have relatively favorable operating environments, and relatively good luck, are adjusted upwards. The general form of the correcting equation is:

$$x^A_{mj} = x_{mj} + [\max\{f^m(z_j;\widehat{\beta^m})\} - f^m(z_j;\hat{\beta}^m)\} + [\max\{\hat{v}_{mj}\} - \hat{v}_{mj}],\quad m = 1,2,\ldots,M;\quad j = 1,2,\ldots,N, \tag{5}$$

where x^A_{mj} and x_{mj} are adjusted and original input values, respectively.

The third stage repeats the linear function (1) using the adjusted values for inputs.

3.2.4. Four-stage model

As in the three-stage model, the first stage of the four-stage model also uses the linear function (1).

In the second stage, the total input slacks (radial plus non-radial) are regressed against a set of uncontrollable variables using Tobit regression. In line with the M inputs, M input slack equations are specified in terms of the following form:

$$S_{mj} = f^m(Z_j;\beta^m;\varepsilon^m),\quad m = 1,2,\ldots,M;\quad j = 1,2,\ldots,N, \tag{6}$$

where $S_{mj} = x_{mj} - X_m\lambda$ are the total slacks from stage one in the usage of the mth input for the jth DMU. Z_j is the vector of uncontrollable variables for the jth DMU. β^m is a vector of coefficients, and ε^m is the statistical noise.

$$x^A_{mj} = x_{mj} + [\max\{f^m(z_j;\widehat{\beta^m})\} - f^m(z_j;\hat{\beta}^m)],\quad m = 1,2,\ldots,M;\quad j = 1,2,\ldots,N, \tag{7}$$

where x^A_{mj} and x_{mj} are adjusted and original input data.

Stage four repeats the DEA in the first stage using the adjusted input data.

4. Data and variables

The research sample used covers 221 Chinese coal-fired power plants during 2002. The plants are largely base load plants and hence can be considered as broadly comparable. The total installed capacity of the sample power plants is 106.7 GW, about 40% of the total coal-fired generating capacity of China in 2002. Their total generation is 577.75 TWh, nearly 45% of the total generation from coal-fired generating capacity. The raw data of individual power plants, such as annual electricity generation, installed capacity, unit scale, number of employees, annual fuel consumption, quality of coal, and vintage, were collected for the sample power plants through fieldwork completed in China by one of the authors between 2005 and 2006. Descriptive statistics of sample power plants are presented in Table 2.

4.1. Traditional variables

Traditional variables include annual electricity generation (MWh), installed capacity (MW), labour (number of employees), and fuel consumption (energy input, TJ). Descriptive statistics for all of these traditional variables are presented in Table 2.

[3] We need to make an assumption about the form of f^m. Because we have no prior knowledge about this and also because several of our variables are categorical, we assume a simple linear functional form.

Note that in this study fuel consumption is measured in terms of energy input (or heat input). This is because in almost all Chinese 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 varies greatly in terms of type of boiler and design of combustion facilities. Generally speaking, given the certain load of a boiler, the more oil it burns, the less coal it consumes. So in order to make the final efficiency evaluation accurate, all fuel consumption, such as coal, oil, and gas, is converted to energy input and is measured in TJ.

4.2. Undesirable output

Undesirable outputs of coal-fired power plants refer to those emissions jointly produced with electricity. Emissions from coal combustion mainly comprise CO_2, SO_2, CH_4, N_2O, NO_x, CO, and non-methane volatile organic compounds (NMVOC). Of these emissions, SO_2 attracts major attention at present. In 2005 the total for emissions in China was about 25.49 million tonnes and this ranked China as number one in the world (SEPA, 2006). Given its importance and the availability of data, SO_2 emissions are selected as the undesirable output in this research. An accurate estimation of SO_2 emissions depends on having knowledge of combustion conditions, technology, and emission control policies, as well as fuel characteristics. The method of estimation for SO_2 emissions is derived from the IPCC reference approach. The general formula for estimating emissions can be described using the following formula:

$$SO_2 \text{ emissions} = 2 \times S_{\text{content}} \times \left(\frac{100 - r}{100}\right) \times \left(\frac{100 - n}{100}\right),$$

where, 2 is the ratio of molecular weight of SO_2 to S(kg/kg); S_{content} is the sulphur content in fuel (tonne); r is the rate of retention of sulphur in ash (%); n is the efficiency of desulphurization (%).

Sulphur content is estimated in terms of annual fuel consumption and coal quality factors. Descriptive statistics for SO_2 emissions are presented in Table 2.

4.3. Uncontrollable variables

Because a wide range of variables may influence efficiency, to limit possibility this research focuses on four variables, including vintage of generating units, calorific value of coal, unit scale and combined heat & power (CHP). Among which vintage is a time-dependent variable, calorific value of coal is directly related to the geological features of coal mines, and unit scale and CHP are characteristics of generating units made at the design and construction stage and they both cannot be changed in operation without remaking the whole thermal system of a power plant. Therefore, it would be right to say that these four variables are independent of the control of management in this occasion.

4.3.1. Vintage of generating units

There are two reasons which support the selection of vintage of generating units. Above all, it is expected that the performance of generating units with regards to both heat rate and availability decreases as the units age (Joskow and Schmalensee, 1987)[4]. Secondly, although there is no definite trend in the efficiency change of auxiliary equipment with time, it is expected to decrease because of wear and tear on the equipment. Additionally, following Pollitt (1995), this paper also tests the impact of variable vintage squared.

4.3.2. Calorific value of coal

The quality of coal affects the operating performance of a coal-fired generating unit. C*eteris paribus,* coal with a higher calorific value tends to be lower in ash content and other impurities, both of which can cause the excessive loading of coal mills and draught fans in the boiler and electrostatic precipitators. Generally, as the calorific value of coal falls, the amount of coal consumed increases and the probability of outages and unit derating also increases (Joskow and Schmalensee, 1987).

4.3.3. Unit scale

Modern technology has made it possible to build generating units with higher parameters, e.g. steam pressure and temperature. At least in theory, increases in both parameters lead to an improvement in thermal efficiency. However, because larger units usually have poorer availabilities than smaller units, the advantage of larger units for thermal efficiency may disappear when the costs of poor reliability are factored in Joskow and Schmalensee (1987). Therefore, it is necessary to check the effects of unit scale on unit performance. Three dummy variables are used in this research, including scale 1 ($0 < \text{scale} < 200$ MW), scale 2 ($200\text{ MW} \leqslant \text{scale} < 400$ MW) and scale 3 ($400\text{ MW} \leqslant \text{scale}$), to test the hypothesis that larger generating units enjoy a higher performance than smaller units. Because scale 2 is the largest category in the sample (Table 2), it is therefore selected as the reference category in the following regression analysis.

4.3.4. Combined heat & power (CHP)

Combined heat & power (CHP), also known as cogeneration, is a type of generating facility which produces electricity and heat (or steam) for industrial, commercial, heating, or for other purposes. Reported fuel consumption for CHP facilities has already been downwardly adjusted by the data providers to account for fuel used for heat. The nature of this adjustment is expected to bias the final efficiency score upwards. Therefore, in order to accurately specify a power plant's efficiency in this research a dummy variable CHP is created, which is equal to 1 if a power plant installs a CHP facility and 0 otherwise.

5. Results

We estimate six models based on the discussion in Section 3 and these are summarized in Table 3. The basic DEA model (model 1) is used as the starting model, which only considers the effects of the undesirable output. Based on this basic model, other models are then used to examine the impact of the operating environment. The calculation of the basic DEA model is performed using the common DEA program DEAP Version 2.1 (Coelli, 1996a). The summary statistics for the results of the basic model are presented in Table 9.

5.1. One-stage model (model 2)

Since the DEA linear function cannot accommodate dummy variables, only two uncontrollable variables, such as vintage of generating units and the calorific value of coal, are included; of which the calorific value of coal is incorporated as a variable with a positive impact and vintage as a variable with a negative influence. The rationality of this arbitrary decision regarding the influence direction of both variables is validated by the later models. Since the program DEAP cannot accommodate the uncontrollable variables, the calculation of this model is carried out by a Matlab

[4] Heat rate refers to the amount of fuel used to generate a KW h of electricity. It can be gross or net heat rate. The difference between these two accounts for the electricity consumed within the power plant itself to run auxiliary equipment.

Table 3
Overview of models and variables

Model name	Basic model (1)	One-stage model (2)	Two-stage model Tobit (3)	Two-stage model logistic (4)	Three-stage model (5)	Four-stage model (6)
Desirable output:						
Annual generation	√	√	√	√	√	√
Inputs:						
Installed capacity	√	√	√	√	√	√
Labour	√	√	√	√	√	√
Fuel	√	√	√	√	√	√
Undesirable output:						
SO_2 emissions	√	√	√	√	√	√
Uncontrollable variables:						
Vintage		√			√	√
Vintage squared			√	√	√	√
Calorific value of coal		√	√	√	√	√
Scale 1			√	√	√	√
Scale 2			√	√	√	√
Scale 3			√	√	√	√
CHP			√	√	√	√

program written by the authors. The summary statistics of efficiency scores for the one-stage model are presented in Table 9.

5.2. Two-stage model

Based on the basic model, in the second stage of the two-stage model efficiency scores from the first-stage are regressed against the selected uncontrollable variables. Both the Tobit model (model 3) and the logistic regression model (model 4) are demonstrated here. The initial results of both models are presented in Table 4.

Except for the coefficient estimated for vintage, in Table 4 the coefficients for other variables are significantly different from zero at 5% or better significance levels in at least one model. Both the *T*-test and log-likelihood ratio test do not reject the hypothesis that vintage makes no contribution to the fitness of the model. Therefore, vintage is discarded in both models and the final results of the two-stage models are presented in Table 5.

Compared with those in Table 4, the coefficients estimated in Table 5 have largely been improved in quality. Table 5 reveals some very important findings. Above all, the variables selected can provide a relatively good explanation of the variation of the efficiency scores. Except for the coefficient for CHP in the logistic regression model, other estimated figures are different from zero at a 5% or better significance level in both models. Secondly, the signs of the estimated coefficients for uncontrollable variables are quite stable in both models and are consistent in an engineering sense. This confirms the impact of the selected uncontrollable variables and their influence direction on the technical efficiency of power plants. For example, vintage does matter for the efficiency variation of a coal-fired power plant, but it is most likely to influence the performance of coal-fired power plants in a negative quadratic way. Therefore, the hypothesis that operating environment has no effect on efficiency scores can be rejected.

The summary statistics for the efficiency scores of this model are presented in Table 9. Due to the adjustment of the final scores for uncontrollable variables, the maximum value predicted by the Tobit model is slightly different from one. This occurs because a relatively efficient power plant is operating in a relatively unfavorable operating environment. Regressions in both models are performed using the statistical package STATA.

Table 4
Initial results of two-stage models

Independent variables	Tobit		Logistic regression	
	Coefficients	Standard error	Coefficients	Standard error
Constant	0.7738^{***}	0.0448	−0.0429	1.7642
Vintage	0.0020	0.0017	−0.0032	0.0676
Vintage squared	-0.00013^{*}	0.00005	−0.0016	0.0021
Calorific value of coal	0.0056^{**}	0.0019	0.1471^{*}	0.0732
Scale 1	-0.0633^{***}	0.0101	-0.8239^{*}	0.3991
Scale 2	–		–	
Scale 3	0.0771^{***}	0.0192	2.3382^{***}	0.7269
CHP	0.0399^{**}	0.0131	0.5965	0.5140
Log-likelihood function	237.9844		−526.0825	

Note: Dependent variables are efficiency scores from the first stage DEA model. ***, ** and * indicate that the parameter estimate is significantly different from zero at the 0.1%, 1% and 5% levels, respectively.

Table 5
Final results of two-stage models

Independent variables	Tobit		Logistic regression	
	Coefficients	Standard error	Coefficients	Standard error
Constant	0.79079^{***}	0.04266	−0.06805	1.67338
Vintage squared	-0.00008^{***}	0.00002	-0.00166^{*}	0.00080
Calorific value of coal	0.00533^{**}	0.00186	0.14748^{*}	0.07272
Scale 1	-0.06377^{***}	0.01015	-0.82326^{*}	0.39885
Scale 2	–		–	
Scale 3	0.07440^{***}	0.01910	2.34216^{***}	0.72193
CHP	0.04190^{***}	0.01299	0.59336	0.50954
Log-likelihood function	237.28035		−526.08358	

Note: Dependent variables are efficiency scores from the first stage DEA model. ***, ** and * indicate that the parameter estimate is significantly different from zero at the 0.1%, 1% and 5% levels, respectively.

5.3. Three-stage model

The first stage of the three-stage model is also the basic DEA model (1). In the second stage, since slacks cannot be aggregated across non-commensurate variables, four SFA slack regressions are conducted. The total slacks (radial plus non-radial) are regressed on the selected uncontrollable variables, based on a truncated normal specification of the one-sided inefficiency error component $u_{mj} \sim N^{+}(\mu^{m}, \sigma^{2}_{um})$[5]. The initial stochastic frontier estimation results are summarized in Table 6.

[5] Because the magnitude of the total slacks might correlate to the size of a power plant, therefore, the total slacks are expressed as percentages both here and later in the four-stage model.

Table 6
Initial stochastic frontier estimation results

Independent variables	Dependent variables			
	Capacity slack	Labour slack	Fuel slack	SO_2 slack
Constant	0.4635*** (0.1043)	1.0780*** (0.1116)	0.2251** (0.0764)	0.2858*** (0.0503)
Vintage	−0.0101* (0.0040)	−0.0030 (0.0050)	−0.0031 (0.0016)	−0.0036 (0.0020)
Vintage squared	0.0003* (0.0001)	0.0001 (0.0002)	0.00018*** (0.00005)	0.00017** (0.00006)
Calorific value of coal	−0.0089* (0.0040)	−0.0063 (0.0044)	−0.0051** (0.0018)	−0.0072** (0.0022)
Scale 1	0.0354 (0.0231)	0.0614* (0.0293)	0.0646*** (0.0097)	0.0554*** (0.0121)
Scale 2	–	–	–	–
Scale 3	−0.1128* (0.0435)	−0.2504*** (0.0570)	−0.0658*** (0.0179)	−0.0767*** (0.0218)
CHP	−0.0900** (0.0299)	−0.0099 (0.0323)	−0.0419** (0.0133)	−0.0398* (0.0165)
σ^2	0.2447*** (0.0020)	0.1258*** (0.0351)	0.0040*** (0.0009)	0.0060*** (0.0007)
γ	0.00003 (0.01165)	0.99999*** (0.00001)	0.0092 (0.2543)	0.0020 (0.0355)
μ	0.0017 (0.0233)	0.3347*** (0.0747)	−0.0120 (0.0653)	−0.0069 (0.0509)
Log-likelihood function	96.4228	−5.9358	298.5798	251.8603

Note: ***, ** and * indicate that the parameter estimate is significantly different from zero at the 0.1%, 1% and 5% levels, respectively.

Table 7
Final estimation results of three-stage model

Independent variables	Dependent variables			
	Capacity slack	Labour slack	Fuel slack	SO_2 slack
Constant	−0.4618*** (0.1072)	1.0780*** (0.1116)	0.2223*** (0.0429)	0.2845*** (0.0531)
Vintage	−0.0101* (0.0041)	−0.0030 (0.0050)	−0.0031 (0.0016)	−0.0036 (0.0020)
Vintage squared	0.0003* (0.0001)	0.0001 (0.0002)	0.00018*** (0.00005)	0.0002** (0.0001)
Calorific value of coal	−0.0089* (0.0044)	−0.0063 (0.0044)	−0.0051** (0.0018)	−0.0072**(0.0022)
Scale 1	0.0354 (0.0242)	0.0614* (0.0293)	0.0647*** (0.0097)	0.0554*** (0.0120)
Scale 2	–	–	–	–
Scale 3	−0.1128** (0.0442)	−0.2504*** (0.0570)	−0.0659*** (0.0177)	−0.0766*** (0.0219)
CHP	−0.090** (0.0312)	−0.0099 (0.0323)	−0.0421*** (0.0125)	−0.0398* (0.0155)
σ^2	0.0253	0.1258	0.0041	0.0062
γ	–	0.99999***	–	–
μ	–	0.3347***	–	–
Log-likelihood function	96.4228	−5.9358	298.5805	251.8609

Note: ***, ** and * indicate that the parameter estimate is significantly different from zero at the 0.1%, 1% and 5% levels, respectively.

A likelihood ratio test does not reject the hypothesis that the $u_{mj} \sim N^{+}(\mu^{m}, \sigma^{2}_{um})$ makes no contribution to the composed error term $(v_{mj} + u_{mj})$, except for with the labour slack regression. Also, a *T*-test on the $\gamma^{m} = \sigma^{2}_{um}/(\sigma^{2}_{vm} + \sigma^{2}_{um})$ shows the hypothesis that $\gamma^{m} = 0$ cannot be rejected at any sensible importance level in the regressions for capacity slack, fuel slack and SO_2 slack. Therefore, both tests disclose that managerial inefficiency actually exerts no impact on the usage of these three inputs.

On the contrary, the contribution of managerial inefficiency to the labour slack is confirmed by both a likelihood ratio test and a *T*-test. Table 6 shows that the estimated value of γ in the SFA regression for labour slack is very close to one[6]. That is to say, managerial inefficiency is actually able to explain almost all variations in the labour slack.

Given the aforementioned reasons, the SFA regressions for capacity, fuel and SO_2 slacks are replaced by three OLS regressions[7]. The new estimation results are presented in Table 7.

Based on the results in Table 7, the input data is adjusted and the DEA model is repeated. The summary statistics of the efficiency scores of the three-stage model are presented in Table 9. The SFA regression is implemented using the program FRONTIER, Version 4.1 (Coelli, 1996b). Input data are adjusted in Microsoft EXCEL. Final efficiency scores are calculated using the program DEAP, Version 2.1.

5.4. Four-stage model

As mentioned previously, the procedure for the four-stage model is quite similar to that of the three-stage model. Based on the Tobit regression technique, the estimated coefficients of the slack regressions are listed in Table 8.

It can be seen that the signs of the coefficients are fairly consistent. The coefficients estimated for the selected uncontrollable variables are significantly important. The results again demonstrate the impact of the operating environment on a power plant's performance. The input data are then adjusted and the DEA model is repeated. Its summary statistics are also presented in Table 9. The four Tobit regressions are implemented using the statistical package STATA. Input data are adjusted in Microsoft EXCEL. The DEA calculation in the last stage is performed using the program DEAP, Version 2.1.

5.5. Model comparison

Table 9 shows the summary statistics of the efficiency scores for different models. Although they clearly do differ, after the effects of the uncontrollable variables are considered there is a general increase in efficiency scores across the models, from models (2)–(6). This suggests that the impact of the uncontrollable variables on the efficiency of coal-fired power plants is fairly stable.

[6] These kinds of very high values for γ are also reported by other researchers, e.g. Jamasb and Pollitt (2003) and Coelli and Perelman (1996).

[7] We note in passing that in the original implementation of the three-stage model, Fried et al. (2002) made use of Tobit regression in a similar situation. However, we think that it is logically more consistent to use OLS, given that we started with linear SFA. The difference this makes to the final results is negligible.

Table 8
Estimation results of four-stage model

Independent variables	Dependent variables			
	Capacity slack	Labour slack	Fuel slack	SO_2 slack
Constant	−0.4809*** (0.1146)	0.5982** (0.1988)	0.2303*** (0.0460)	0.2956*** (0.0567)
Vintage	−0.0101* (0.0044)	0.0069 (0.0076)	−0.0031 (0.0018)	−0.0036 (0.0022)
Vintage squared	0.0003* (0.0001)	−0.0001 (0.0002)	0.00017*** (0.00005)	0.0002** (0.0001)
Calorific value of coal	−0.0101* (0.0048)	−0.0080 (0.0082)	−0.0056** (0.0019)	−0.0079*** (0.0024)
Scale 1	0.0378 (0.0259)	0.1097* (0.0449)	0.0661*** (0.0104)	0.0571*** (0.0127)
Scale 2	–	–	–	–
Scale 3	−0.1355** (0.0487)	−0.3121*** (0.0843)	−0.0763*** (0.0197)	−0.0904*** (0.0243)
CHP	−0.0940** (0.0334)	−0.0734 (0.0579)	−0.0433*** (0.0134)	−0.0412* (0.0165)
σ^2	0.1685*** (0.0086)	0.2929*** (0.0150)	0.0677*** (0.0034)	0.0834*** (0.0042)
Log-likelihood function	51.5552	−62.4791	233.0675	192.6340

Note: ***, ** and * indicate that the parameter estimate is significantly different from zero at the 0.1%, 1% and 5% levels, respectively.

Table 9
Summary statistics of efficiency scores

Model name	Basic model (1)	One-stage model (2)	Two-stage model Tobit (3)	Two-stage model logistic (4)	Three-stage model (5)	Four-stage model (6)	Model average (2)–(6)
Mean	0.884	0.910	0.884	0.889	0.950	0.901	0.907
Standard error	0.076	0.069	0.062	0.083	0.036	0.066	0.053
Minimum	0.648	0.664	0.684	0.576	0.799	0.672	0.724
Maximum	1	1	1.058	1	1	1	1.012
Number of efficient DMU	20	41	6	20	20	18	–

Table 10
Average efficiency scores in terms of vintage

Vintage (year)	No. of power plants	Basic model (1)	One-stage model (2)	Two-stage model Tobit (3)	Two-stage model logistic (4)	Three-stage model (5)	Four-stage model (6)	Model average (2)–(6)
0–10	136	0.894	0.904	0.881	0.876	0.955	0.910	0.905
11–20	58	0.884	0.907	0.887	0.897	0.950	0.892	0.906
21–30	21	0.852	0.943	0.903	0.935	0.933	0.886	0.920
31–	6	0.766	0.946	0.866	0.962	0.875	0.816	0.893

A Rank–Sum test[8] is implemented in order to identify whether or not the difference is significant between the efficiency scores of the basic model (1) and those averages of the models (2)–(6). The Rank–Sum statistic achieved is about 3.257. It indicates that the null hypothesis that the two groups of scores belonging to the same distribution can be rejected at the significance level of 0.5%. This demonstrates that both groups of efficiency scores are significantly different. In other words, incorporating the selected uncontrollable variables does make a difference in the final efficiency evaluation.

Some results related to Table 9 are also worthy of stressing here. First of all, DMUs operating in relatively unfavourable environment tends to have larger efficiency score differences between the basic model and other models. Remember that our attempt to eliminate the effects of uncontrollable variables is to level the playing ground for all DMUs. In the one-stage model, this is realized by comparing the *j*th DMU with a theoretical frontier DMU operating in an environment that is no better than that of the *j*th DMU. In the two-stage model, this is done by adjusting each DMU's efficiency score to the average level of the estimated impact of uncontrollable variables. In the three- and four-stage models, this is done by adjusting upward the inputs of those DMUs, who have relatively favourable operating environments. In general, the bigger the adjustment is, the larger the efficiency score difference between the basic model and other models will be. The final results support this argument. Secondly, except in the one-stage model and the two-stage Tobit model, efficient DMUs in other models are quite consistent. The one-stage model has the largest number of efficient DMUs – this agrees with what the authors discussed in Section 2.2.2. That is, as variables included in DEA increase, the number of efficient DMUs will increase. Thirdly, when compared with other models, the three-stage model has the narrowest efficiency score interval (Table 9). This to some extent reflects the superiority of the three-stage model to eliminate the impact of uncontrollable variables.

Tables 10–13 exhibit the relationships between the average efficiency scores and the selected uncontrollable variables, these being vintage, calorific value of coal, unit scale, and CHP. When compared with the average efficiency scores of the basic model (1), those of models (2)–(6) become less tendentious after being adjusted by the corresponding uncontrollable variables. This result validates the authors' selection of uncontrollable variables in the previous section. It is also consistent with the hypothesis that at least some of the power plants which had relatively low efficiency scores in the basic model, did so in part due to their relatively unfavorable operating environments. Not all of them were as poorly managed as their low efficiency scores in the traditional DEA model indicated.

Tables 14 and 15 exhibit the simple and the rank correlations of the efficiency scores respectively. Generally, a high correlation coefficient between two sets of efficiency scores indicates a high

[8] The Rank–Sum test, also known as the Wilcoxon–Mann–Whitney test, is a kind of nonparametric statistic. Since the theoretical distribution of the efficiency score in DEA is usually unknown, therefore, using the parametric approach in this context is more susceptible. Please see Brockett and Golany (1996) and Cooper et al. (2000) for detail.

Table 11
Average efficiency scores in terms of calorific value of fuel

Calorific value (TJ/tonne)	No. of power plants	Basic model (1)	One-stage model (2)	Two-stage model Tobit (3)	Two-stage model logistic (4)	Three-stage model (5)	Four-stage model (6)	Model average (2)–(6)
<20	23	0.859	0.879	0.884	0.907	0.941	0.898	0.902
20–25	164	0.885	0.911	0.886	0.894	0.950	0.902	0.909
25<	34	0.900	0.924	0.874	0.852	0.954	0.896	0.900

Table 12
Average efficiency scores in terms of the scale dummy variable

Scale	No. of power plants	Basic model (1)	One-stage model (2)	Two-stage model Tobit (3)	Two-stage model logistic (4)	Three-stage model (5)	Four-stage model (6)	Model average (2)–(6)
Scale 1	91	0.841	0.885	0.886	0.908	0.927	0.877	0.897
Scale 2	115	0.906	0.920	0.884	0.889	0.962	0.915	0.914
Scale 3	15	0.974	0.981	0.873	0.778	0.990	0.933	0.911

Table 13
Average efficiency scores in terms of the CHP dummy variable

CHP	No. of power plants	Basic model (1)	One-stage model (2)	Two-stage model Tobit (3)	Two-stage model logistic (4)	Three-stage model (5)	Four-stage model (6)	Model average (2)–(6)
CHP (=1)	46	0.874	0.936	0.886	0.903	0.941	0.875	0.908
CHP (=0)	175	0.887	0.903	0.884	0.886	0.952	0.907	0.906

Table 14
Correlation of efficiency scores for different models

	Basic	One-stage	Two-stage (Tobit)	Two-stage (logistic)	Three-stage	Four-stage
Basic	1.000					
One-stage	0.775	1.000				
Two-stage (Tobit)	0.775	0.686	1.000			
Two-stage (logistic)	0.311	0.374	0.701	1.000		
Three-stage	0.980	0.701	0.731	0.268	1.000	
Four-stage	0.885	0.657	0.863	0.497	0.893	1.000

Table 15
Spearman's rank correlation of efficiency scores

	Basic	One-stage	Two-stage (Tobit)	Two-stage (logistic)	Three-stage	Four-stage
Basic	1.000					
One-stage	0.792	1.000				
Two-stage (Tobit)	0.738	0.667	1.000			
Two-stage (logistic)	0.441	0.520	0.819	1.000		
Three-stage	0.988	0.765	0.694	0.401	1.000	
Four-stage	0.869	0.671	0.835	0.601	0.875	1.000

consistency in both sets. It is clear to see that the efficiency scores of the three-stage and four-stage models have, in general, a higher correlation with the other models[9]. It indicates that these two models are able to explain most of the features of the other models, thus suggesting their superiority. Relative to the four-stage model, the three-stage model is able to differentiate between managerial inefficiency and statistical noise. This feature is valuable for an understanding of the causes of DMU inefficiency. Therefore, it is methodologically more preferable.

6. Conclusion

There are very few published studies on performance measurement which simultaneously incorporate both undesirable outputs and uncontrollable variables. For a few papers in which undesirable outputs have been considered, the impact of uncontrollable variables has been left unexamined. That is to say, many existing studies on performance measurement implicitly assume that all inefficiency is due to the bad management of DMUs. Therefore, the performance of DMUs in a relatively unfavorable operating environment is very likely to be underestimated.

In this research different DEA-based efficiency measurement models are used to examine the impact of uncontrollable variables together with undesirable outputs. Furthermore, the basic model is derived from existing literature and is only able to consider the impact of undesirable outputs. The one-stage model incorporates selected uncontrollable variables directly into the DEA model. Despite being simple, it requires a prior understanding of the direction of uncontrollable variables. Also, it cannot accommodate uncontrollable dummy variables. The two-stage model involves constructing a basic DEA model for considering the impact of undesirable outputs in the first stage. The radial efficiency scores are then regressed against a set of uncontrollable variables in the second stage. Its advantage is that this model does not require any prior knowledge of the influence direction of uncontrollable variables. However, the two-stage model ignores the information contained in the input slacks. Moreover, because DEA is an extreme point method, any outlier is likely to influence the position of the efficiency frontier. Therefore, the use of efficiency scores gathered from the first stage as a dependent variable in the second stage does in theory entail an amount of risk. When compared with the two-stage model, the three-stage and four-stage models can

[9] As one of the reviewers pointed out, the correlations between the efficiency scores in the various situations might be influenced by the truncated errors or the fact that many DMUs have efficiency scores equal to one. However as Table 9 indicates the number of fully efficient DMUs is quite small as a percentage of the total sample size. Nor do the differences in correlation coefficient between the methods seem to vary with number of fully efficient DMUs, in particular models 2 and 3 with the most different number of efficient DMUs do not seem to be unusual with reasonably high cross correlation to each other and similar cross correlations to the other models.

make better use of any information contained in the input slacks. Our research results confirm their superiority over other models. They are both demanding in computation and the three-stage model. The three-stage, which uses the SFA technique as a regression tool, is able to differentiate between managerial inefficiency and statistical noise. It can also give a better explanation of DMU inefficiency. It is therefore more preferable.

This paper not only contributes to performance measurement research methodology, but it also has implications for policies affecting the Chinese coal-fired power sector. With the rapid increase of coal-fired generating capacity, which has been a dominant part of Chinese electricity generation for decades, emissions from electricity generation have caused enormous environmental damage and great social economic cost in China. Given the large scale of the sector and its worldwide importance with regards to the control of climate change, how to raise the efficiency of coal-fired power plants has become a crucial issue. 221 coal-fired power plants are pooled in a research sample in this paper, along with data on their annual generation, capital, labour, and fuel consumption, and also with data on their vintage, calorific value of coal and unit scale. The results obtained indicate that the impact of uncontrollable variables is relatively significant. This confirms the hypothesis that at least some power plants with relatively low efficiency scores in the traditional model achieved these in part due to their relatively unfavorable operating environments. However, it should be noted that after correcting this we find inefficiency to be around 10%. Eliminating this inefficiency via the appropriate market and regulatory mechanisms would yield substantial economic and environmental benefits. Future research will examine power plant efficiency in more detail and incorporate additional undesirable outputs, such as CO_2.

References

Banker, R., Morey, R., 1986. Efficiency analysis for exogenously fixed inputs and outputs. Operation Research 34 (4).

Banker, R., Kaufman, R., Morey, R., 1990. Measuring gains in operational efficiency from information technology: A study of the positran deployment at Hardees Inc. Journal of Management Information System 7 (2), 29–54.

Brockett, P., Golany, B., 1996. Using rank statistics for determining programmatic efficiency differences in data envelopment analysis. Management Science 42 (3), 466–472.

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

Charnes, A., Cooper, W., Rhodes, E., 1981. Evaluating program and managerial efficiency: An application of data envelopment analysis to program follow through. Management Science 27 (6).

Coelli, T., 1996a. A guide to DEAP version 2.1: A data envelopment analysis (computer) program. Centre for Efficiency and Productivity Analysis, Department of Econometrics, University of New England, Australia.

Coelli, T., 1996b. A guide to FRONTIER version 4.1: A computer program for stochastic frontier production and cost function estimations. Centre for Efficiency and Productivity Analysis, Department of Econometrics, University of New England, Australia.

Coelli, T., 1997. Total factor productivity growth in Australian coal-fired electricity generation: A malmquist index approach. In: Paper Presented at the International Conference on Public Sector Efficiency, 27–28 November, UNSW, Sydney.

Coelli, T., Perelman, S., 1996. Efficiency measurement, multiple-output technologies and distance functions: With application to European railways. CREPP, Faculty of Economics, University of Liege.

Coelli, T., Rao, D., O'Donnell, C., Battese, G., 2005. An introduction to efficiency and productivity analysis, second ed. Springer, US.

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

Färe, R., Grosskopf, S., Logan, J., Lovell, C., 1985. Measuring efficiency in production: With an application to electric utilities. In: Dogramaci, A., Adam, N. (Eds.), Managerial Issues in Productivity Analysis. Kluwer-Nijhoff, Boston.

Fare, R., Grosskopf, S., Lovell, K., 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., 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.

Fizel, J., Nunnikhoven, T., 1992. Technical efficiency of for-profit and non-profit nursing homes. Managerial and Decision Economics 13, 429–439.

Fried, H., Lovell, C., Eeckaut, P., 1993. Evaluating the performance of US credit unions. Journal of Banking and Finance 17 (2/3), 251–265.

Fried, H., Schmidt, S., Yaisawarng, S., 1999. Incorporating the operating environment into a nonparametric measure of technical efficiency. Journal of Productivity Analysis 12 (3), 249–267.

Fried, H., Lovell, C., Schmidt, S., Yaisawarng, S., 2002. Accounting for environmental effects and statistical noise in data envelopment analysis. Journal of Productivity Analysis 17, 157–174.

Grosskopf, S., Valdmanis, V., 1987. Measuring hospital performance: A non-parametric approach. Journal of Health Economics 6, 89–107.

Jamasb, T., Pollitt, M., 2003. International benchmarking and regulation: An application to European electricity distribution utilities. Energy Policy 31, 1609–1622.

Joskow, P., Schmalensee, R., 1987. The performance of coal-burning electric generating units in The United States: 1960–1980. Journal of Applied Econometrics 2, 85–109.

Korhonen, P., Luptacik, M., 2004. Eco-efficiency analysis of power plants: An extension of data envelopment analysis. European Journal of Operational Research 154, 437–446.

McCarty, T., Yaisawarng, S., 1993. Technical efficiency in New Jersey school districts. In: Fried, H., Lovell, C., Schmidt, S. (Eds.), The Measurement of Productive Efficiency: Techniques and Applications. Oxford Unviersity Press, New York.

Olatubi, W., Dismukes, D., 2000. A data envelopment analysis of the levels and determinants of coal-fired electric power generation performance. Utilities Policy 9, 47–59.

Pastor, J., 2002. Credit risk and efficiency in the European banking system: A three-stage analysis. Applied Financial Economics 12, 895–911.

Pittman, R., 1983. Multilateral productivity comparisons with undesirable outputs. The Economic Journal 93, 883–891.

Pollitt, M.G., 1995. Ownership and Performance in Electric Utilities. Oxford University Press.

Simar, L., Wilson, P., 2007. Estimation and inference in two-stage, semi-parametric models of production processes. Journal of Econometrics 136, 31–64.

State Environmental Protection Administration of China (SEPAC). China Forum of Environmental Journalists. Available from: <http://www.cfej.net/news> (accessed 08.08.06.).

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.

Yaisawarng, S., 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.

Yang, H., Pollitt, M., 2007. Distinguish weak and strong disposability among undesirable outputs in DEA: The example of the environmental efficiency of chinese coal-fired power plants. Working Paper No. CWPE0741 and EPRG0717, Cambridge University, UK.

Zhang, A., 2006. Analysis on the energy efficiency of Chinese electricity. China Energy, 7 (Beijing, in Chinese).