Centre for Efficiency and Productivity Analysis

Working Paper Series
No. WP03/2025

Accounting for Environmental Variables
and Bad Outputs in DEA

C.J. O'Donnell

Date: September 2025

School of Economics
University of Queensland
St. Lucia, Qld. 4072
Australia

ISSN No. 1932 - 4398

Accounting for Environmental Variables and Bad Outputs in DEA

C.J. O'Donnell
University of Queensland
c.odonnell@economics.uq.edu.au

12 September, 2025

Abstract

Most data envelopment analysis (DEA) estimators ignore environmental variables and/or the production of bad outputs. This leads to biased and inconsistent estimates of production frontiers and associated measures of firm performance. This paper develops DEA estimators that account for both environmental variables and the production of bads. These estimators provide more accurate and policy-relevant estimates of economic quantities of interest.

1 Introduction

There is a substantial literature on data envelopment analysis (DEA) methods for estimating production frontiers. Very few of these methods distinguish between variables that are within the control of firm managers and those that are not; see, for example, the seminal papers by Charnes et al (1978) and Banker et al (1984). Exceptions include the methods described by Banker and Morey (1986) and O'Donnell (2018). Following O'Donnell (2018), I define an environmental variable to be one that is physically involved in the production process but never within the control of managers (e.g., ambient temperature). DEA estimators that ignore relevant environmental variables yield biased and inconsistent estimates of production frontiers and associated measures of firm performance (e.g., technical efficiency).

In recent years, the performance of firms that produce bad outputs (e.g., greenhouse gas emissions) has become an important area of empirical research. Unfortunately, most DEA methods also ignore the production of bads. Exceptions include the methods described by Färe et al (1989), Ball et al (1994), Hailu and Veeman (2001), Seiford and Zhu (2002) and Kuosmanen (2005). As with environmental variables, DEA estimators that ignore relevant bad outputs yield biased and inconsistent estimates of production frontiers and measures of efficiency.

This paper develops DEA estimators of production frontiers that take explicit account of environmental variables and the joint production of undesirable outputs. Section 2 presents the assumptions that underpin DEA estimation in such settings. Section 3 introduces the proposed estimators and illustrates their use with examples. Section 4 connects the contribution to existing work in the DEA literature. Section 5 discusses implications for empirical applications and outlines avenues for future research.

2 Assumptions

DEA estimators are underpinned by assumptions about what can and cannot be produced using available production technologies. The assumptions I make in this paper are designed to reflect a key characteristic of polluting technologies, namely that the production of goods also involves the production of bads. Formally, let g and b denote vectors of good and bad outputs, and let $P^t(x, z)$ denote the set of good and bad output vectors that can be produced using the input vector x and the technologies available in period t when the production environment is characterised by the vector z. In this paper, I assume that:

A1 All inputs, outputs and environmental variables are observed and measured without error.

A2 Production frontiers are locally linear.

A3 $P^t(x, z)$ is bounded.

A4 Production possibilities sets are convex: mathematically, if $(g, b) \in P^t(x, z)$ and $(\bar{g}, \bar{b}) \in P^t(\bar{x}, z)$, then for all $\lambda \in [0, 1]$, $(\lambda g + (1 - \lambda)\bar{g},\ \lambda b + (1 - \lambda)\bar{b}) \in P^t(\lambda x + (1 - \lambda)\bar{x}, z)$.

A5 Inactivity is possible: $(0, 0) \in P^t(x, z)$.

A6 Good outputs are strongly disposable: if $(g, b) \in P^t(x, z)$ and $0 \leq \bar{g} \leq g$, then $(\bar{g}, b) \in P^t(x, z)$.

A7 Environmental variables are strongly disposable: if $(g, b) \in P^t(x, z)$ and $\bar{z} \geq z$, then $(g, b) \in P^t(x, \bar{z})$.

A8 Inputs are strongly disposable: if $(g, b) \in P^t(x, z)$ and $\bar{x} \geq x$, then $(g, b) \in P^t(\bar{x}, z)$.

Assumptions A1 and A2 imply there are no functional form errors or other sources of statistical noise. Assumptions A3 to A8 imply that, among other things, $P^t(x, z)$ can be represented by various distance functions.

3 DEA Estimators

This section introduces DEA estimators for three period-and-environment-specific distance functions: the good-output distance function (GODF), the bad-output distance function (BODF), and the input distance function (IDF). All three functions measure how far an observed input–output combination is from a production frontier, but they differ in the variables held fixed and the quantities that are scaled when measuring distance to the frontier. Consequently, the three distance functions also differ in their policy relevance. Estimation of the GODF is most relevant when the objective is to expand desirable production without increasing associated bad outputs. For example, in South Asia the challenge is to raise food production to meet the needs of a growing population while avoiding additional greenhouse gas emissions. By contrast, estimation of the BODF is most relevant when the objective is to maintain existing levels of good outputs while reducing bad outputs. For instance, in Australia and New Zealand, where livestock output is already more than sufficient for domestic needs, reducing methane and nitrous oxide emissions from ruminant livestock is a growing concern. Finally, estimation of the IDF is most relevant when the objective is to reduce the use of inputs while maintaining given levels of both good and bad outputs. For example, in many agricultural systems there is scope to lower fertiliser and water use while holding yields and emissions fixed. To formalize these ideas and accommodate applications involving panel data, subscripts will henceforth be used to distinguish between firms and time periods. Specifically, x_{it} will be used to denote the input vector of firm i in period t; g_{it} and b_{it} will be used to denote the corresponding good and bad output vectors; and z_{it} will be used to denote the vector of environmental variables faced by firm i in period t.

3.1 The Good-Output Distance Function

The GODF gives the reciprocal of the largest factor by which it is possible to scale up a given good output vector when using given inputs to produce given bad outputs in a given period in a given production environment. For example, if it is possible for a firm to use its inputs to produce the same bad outputs and four times as much of every good output, then the GODF takes the value $1/4 = 0.25$. By construction, the GODF lies in the unit interval. Its value provides a measure of good-output-oriented technical efficiency (GOTE). The GOTE of a firm (or its manager) reflects how well good outputs are maximised while holding inputs, bad outputs and environmental variables fixed.

If assumptions A1 and A2 are true, then the GODF takes the form:

$$D_G^{it}(x_{it}, g_{it}, b_{it}, z_{it}) = \gamma_{it}' g_{it} / (\alpha_{it} + \delta_{it}' z_{it} + \beta_{it}' x_{it} + \rho_{it}' b_{it}) \tag{1}$$

where α_{it} is an unknown scalar and $\gamma_{it} = (\gamma_{1it}, \ldots, \gamma_{Nit})'$, $\delta_{it} = (\delta_{1it}, \ldots, \delta_{Jit})'$, $\beta_{it} = (\beta_{1it}, \ldots, \beta_{Mit})'$ and $\rho_{it} = (\rho_{1it}, \ldots, \rho_{Kit})'$ are unknown vectors. DEA estimation of the GODF involves maximising $D_G^{it}(x_{it}, g_{it}, b_{it}, z_{it})$ subject to constraints that ensure that assumptions A4 to A8 are satisfied. If there are I firms in the dataset and we allow for technical progress, then assumption A4 will be satisfied if and only if $\gamma_{it}' g_{hr} \leq \alpha_{it} + \delta_{it}' z_{hr} + \beta_{it}' x_{hr} + \rho_{it}' b_{hr}$ for all $h \leq I$ and $r \leq t$. Assumption A5 requires that this constraint also holds when the good and bad output vectors are equal to zero, i.e., if and only if $0 \leq \alpha_{it} + \delta_{it}' z_{hr} + \beta_{it}' x_{hr}$ for all $h \leq I$ and $r \leq t$. Assumptions A6 to A8 will be satisfied if $\gamma_{it} \geq 0$, $\delta_{it} \geq 0$ and $\beta_{it} \geq 0$. With all these constraints, the estimation problem becomes:

$$\begin{aligned} \max_{\alpha_{it}, \delta_{it}, \beta_{it}, \rho_{it}, \gamma_{it}} \Big\{ & \gamma_{it}' g_{it} / (\alpha_{it} + \delta_{it}' z_{it} + \beta_{it}' x_{it} + \rho_{it}' b_{it}) : \gamma_{it} \geq 0,\ \delta_{it} \geq 0,\ \beta_{it} \geq 0, \\ & \alpha_{it} + \delta_{it}' z_{hr} + \beta_{it}' x_{hr} + \rho_{it}' b_{hr} - \gamma_{it}' g_{hr} \geq 0 \text{ for all } h \leq I \text{ and } r \leq t, \\ & \alpha_{it} + \delta_{it}' z_{hr} + \beta_{it}' x_{hr} \geq 0 \text{ for all } h \leq I \text{ and } r \leq t \Big\}. \end{aligned} \tag{2}$$

This is a fractional programming problem with an infinite number of solutions. A unique solution can be identified by setting $\gamma_{it}' g_{it} = 1$. With this so-called normalising constraint, problem (2) can be rewritten as

$$\begin{aligned}\min_{\alpha_{it},\delta_{it},\beta_{it},\rho_{it},\gamma_{it}} \Big\{ &\alpha_{it} + \delta_{it}' z_{it} + \beta_{it}' x_{it} + \rho_{it}' b_{it} : \gamma_{it} \geq 0,\ \delta_{it} \geq 0,\ \beta_{it} \geq 0, \gamma_{it}' g_{it} = 1, \\ &\alpha_{it} + \delta_{it}' z_{hr} + \beta_{it}' x_{hr} + \rho_{it}' b_{hr} - \gamma_{it}' g_{hr} \geq 0 \text{ for all } h \leq I \text{ and } r \leq t, \\ &\alpha_{it} + \delta_{it}' z_{hr} + \beta_{it}' x_{hr} \geq 0 \text{ for all } h \leq I \text{ and } r \leq t \Big\}. \end{aligned} \tag{3}$$

This is a linear program (LP). The value of the objective function at the optimum is an estimate of the reciprocal of the GODF (i.e., the reciprocal of GOTE).

Problem (3) is what many researchers would call a primal LP. Every primal LP has dual form with the property that if the primal and its dual both have feasible solutions, then the optimised values of the two objective functions are equal. Many DEA researchers prefer to work with the dual form, which in this case is

$$\begin{aligned}\max_{\mu,\lambda,\tau} \Big\{ \mu :\ &\mu g_{it} \leq \sum_{h=1}^{I}\sum_{r=1}^{t} \lambda_{hr} g_{hr}, \sum_{h=1}^{I}\sum_{r=1}^{t} \lambda_{hr} b_{hr} = b_{it}, \\ &\sum_{h=1}^{I}\sum_{r=1}^{t} (\lambda_{hr} + \tau_{hr}) z_{hr} \leq z_{it},\ \sum_{h=1}^{I}\sum_{r=1}^{t} (\lambda_{hr} + \tau_{hr}) x_{hr} \leq x_{it}, \\ &\sum_{h=1}^{I}\sum_{r=1}^{t} (\lambda_{hr} + \tau_{hr}) = 1, \lambda_{hr} \geq 0 \text{ and } \tau_{hr} \geq 0 \text{ for all } h \text{ and } r \Big\} \end{aligned} \tag{4}$$

where μ is an unknown scalar and $\lambda = (\lambda_{11}, \ldots, \lambda_{IT})'$ and $\tau = (\tau_{11}, \ldots, \tau_{IT})'$ are unknown vectors. This dual formulation highlights that the DEA estimator of the GODF seeks to scale up the good-output vector while holding bad outputs, inputs and environmental variables fixed. In practice, solutions to both the primal and dual LPs can be obtained using standard DEA software packages by simply adding "inactive" variants of the observed firms to the dataset. These inactive firms use the same input vectors and operate in the same production environments as the observed firms, but they produce no outputs. Most DEA software packages also report the values of shadow prices, i.e., the marginal benefits of increasing good outputs and the marginal costs of reducing bad outputs when all other variables are held fixed. In this way, the GODF estimator not only identifies the potential for expanding desirable production, but also provides economically meaningful measures of trade-offs between goods and bads.

To clarify some of these concepts, consider an industry in which firms use a single input to produce one good output and one bad output in a production environment characterised by a single environmental variable. Fabricated data for six hypothetical firms are reported in Table 1. Because all firms use the same amount of input and operate in the same period in the same production environment, their output choices can be depicted as points A to F in Figure 1. In this figure, the DEA frontier is given by the line segments connecting the horizontal axis with points A, B, C and D; this frontier reflects assumptions A1 to A6. DEA estimates of GOTE for each firm are reported in Table 2. Firms A to D all have estimated GOTE scores of 1, indicating that they could not have increased their production of good outputs while holding inputs and bad outputs fixed. By contrast, Firm E has a GOTE score of 0.8, indicating that it could have increased its production of good outputs by a factor of $1/0.8 = 1.25$, and Firm F has a GOTE score of 0.2, indicating that it could have increased its production of good outputs by a factor of $1/0.2 = 5$. Table 2 also reports estimates of the shadow price of the good output (SPG), the shadow price of the bad output (SPB), and the marginal rate of transformation (MRT = SPB/SPG) for each firm. The SPB measures the revenue loss per unit of bad output (e.g., Firm A loses 0.333 units of shadow revenue for each additional unit of the bad). The MRT represents the rate at which the good output can be substituted for the bad output as we move around the frontier. For firms inside the frontier (e.g., E and F), shadow prices and MRTs are not computed at the observed points themselves but at associated projection points on the frontier (e.g., E″ and F″); in the case of the GODF, these projection points are obtained by radially expanding the observed good-output vector while holding inputs and bad outputs fixed. For points A and B, the estimated MRTs are equal to the slope of the line segment AB; for points C, E and F, they are given by the slope of BC; and for point D, by the slope of CD. Importantly, shadow prices—and therefore MRTs—at the corner points (A, B, C and D) are not unique. Although upper and lower bounds on shadow prices (and thus MRTs) at these points can be obtained by solving secondary linear programs, a full treatment is beyond the scope of this paper; for details, see Cooper et al (2007). The shadow prices reported here (and elsewhere in the paper) are the ones computed by the *R* package of Bogetoft and Otto (2015).[1]

As a second example, consider an industry that uses a single input to produce two good outputs and one bad output in a production environment characterised by a single environmental variable. Fabricated data for six hypothetical firms are reported in Table 3. Because all firms use the same amount of input to produce the same amount of bad output, and because they all operate in the same period in the same production environment, their good output choices can be represented by points A to F in Figure 2. In this figure, the DEA estimate of the frontier is given by the line segments joining the two axes with the frontier points B, C and D; this frontier reflects assumptions A1 to A4 and A6. DEA estimates of GOTE for each firm are reported in Table 4. Observe that the estimated GOTE scores of Firms B, C and D are all equal to 1; this indicates that it would not have been possible for these firms to scale up their production of good outputs while holding their inputs and bad outputs fixed. By contrast, the estimated GOTE scores of Firms A and E are 0.5; this indicates that these firms could have doubled their production of good outputs while holding their inputs and bad outputs fixed. Similarly, the estimated GOTE of Firm F is 0.83; this indicates that this firm could have increased its production of good outputs by a factor of $1/0.83 = 6/5 = 1.2$ while holding its inputs and bad outputs fixed. Table 4 also reports estimates of shadow prices and marginal rates of transformation (MRTs). The MRTs are the negatives of the ratios of the shadow prices (i.e., $\text{MRT}_{21} = -\ \text{SPG}_1/\text{SPG}_2$) and represent the amount of good output 2 that needs to be given up in order to produce an extra unit of good output 1: for point A, the MRT is given by the slope of the horizontal line that joins the vertical axis to point B; for points B and E, it is given by the slope of the line segment BC; for point C it is given by the slope of the line segment CD; and for points D and F, it is given by the slope of the vertical line that joins the horizontal axis to point D.

[1] *R* code for all the examples in the paper is presented in the Appendix.

Table 1: Fabricated Data For Example 1

Firm	Period	Input	Good	Bad	Environ.
i	t	x	g	b	z
A	1	1	3	1	1
B	1	1	5	3	1
C	1	1	5	6	1
D	1	1	2	7	1
E	1	1	4	4	1
F	1	1	1	5	1

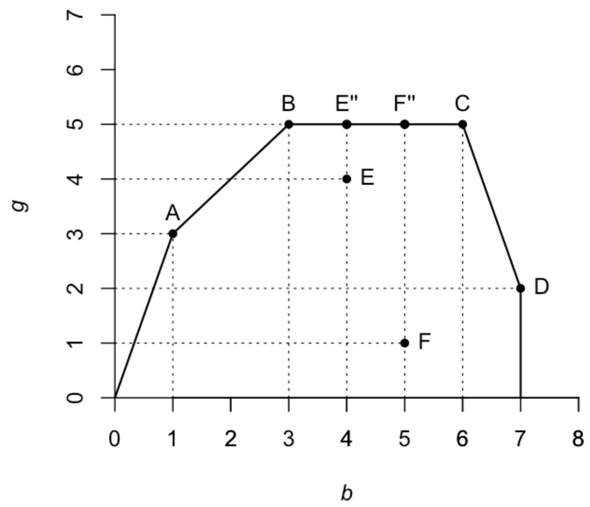

Figure 1: Estimated Output Set For Example 1

Table 2: Economic Quantities of Interest For Example 1

i	GOTE	SPG	SPB	MRT
A	1.0	0.333	0.333	1
B	1.0	0.2	0.2	1
C	1.0	0.2	0.0	0
D	1.0	0.5	−1.5	−3
E	0.8	0.25	0.0	0
F	0.2	1.0	0.0	0

Table 3: Fabricated Data For Example 2

Firm i	Period t	Input x	Good 1 g_1	Good 2 g_2	Bad b	Environ. z
A	1	1	1	3	1	1
B	1	1	3	6	1	1
C	1	1	5	4	1	1
D	1	1	6	2	1	1
E	1	1	2	2.5	1	1
F	1	1	5	1	1	1

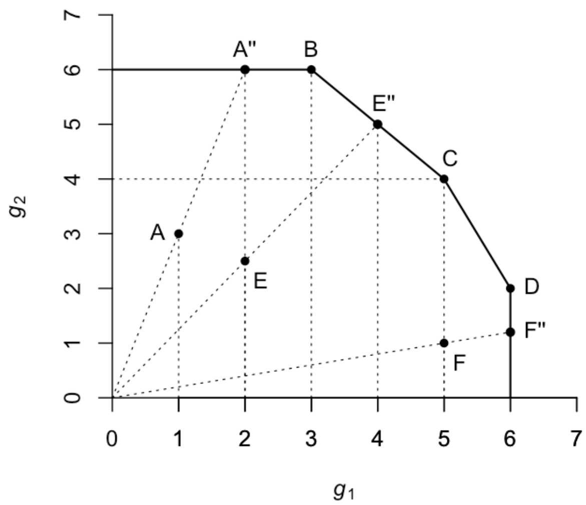

Figure 2: Estimated Output Set For Example 2

Table 4: Economic Quantities of Interest For Example 2

i	GOTE	SPG_1	SPG_2	MRT_{21}
A	0.5	0	0.333	0
B	1	0.111	0.111	−1
C	1	0.143	0.071	−2
D	1	0.167	0.000	Inf
E	0.5	0.222	0.222	−1
F	0.833	0.200	0.000	Inf

3.2 The Bad-Output Distance Function

A period-and-environment-specific BODF gives the reciprocal of the smallest fraction of a given bad output vector that can be produced when using given inputs to produce given good outputs in a given period in a given production environment. For example, if it is possible for a firm to use its inputs to produce its good outputs and as little as one quarter of its bad outputs, then the BODF takes the value $1/0.25 = 4$. By construction, the BODF takes a value no less than one. Its reciprocal lies in the unit interval and provides a measure of bad-output-oriented technical efficiency (BOTE). The BOTE of a firm (or its manager) reflects how well bad outputs are minimised while holding inputs, good outputs and environmental variables fixed.

If assumptions A1 and A2 are true, then the BODF takes the form:

$$D_B^{it}(x_{it}, g_{it}, b_{it}, z_{it}) = (\upsilon_{it}' b_{it})/(\eta_{it}' g_{it} - \psi_{it}' z_{it} - \omega_{it}' x_{it} - \sigma_{it}) \tag{5}$$

where σ_{it} is an unknown scalar and $\eta_{it} = (\eta_{1it}, \ldots, \eta_{Nit})'$, $\psi_{it} = (\psi_{1it}, \ldots, \psi_{Jit})'$, $\upsilon_{it} = (\upsilon_{1it}, \ldots, \upsilon_{Kit})'$ and $\omega_{it} = (\omega_{1it}, \ldots, \omega_{Mit})'$ are unknown vectors. In this case, DEA estimation involves minimising $D_B^{it}(x_{it}, g_{it}, b_{it}, z_{it})$ subject to constraints that ensure that assumptions A4 to A8 are satisfied. Applying the same reasoning as in Section 3.1, the primal LP turns out to be:

$$\begin{aligned} \max_{\xi_{it}, \kappa_{it}, \phi_{it}, \nu_{it}, \theta_{it}} \Big\{ & \eta_{it}' g_{it} - \psi_{it}' z_{it} - \omega_{it}' x_{it} - \sigma_{it} : \eta_{it} \geq 0,\ \psi_{it} \geq 0,\ \omega_{it} \geq 0, \upsilon_{it}' b_{it} = 1, \\ & \eta_{it}' g_{it} - \psi_{it}' z_{it} - \omega_{it}' x_{it} - \sigma_{it} - \upsilon_{it}' b_{it} \leq 0 \text{ for all } h \leq I \text{ and } r \leq t, \\ & - \psi_{it}' z_{it} - \omega_{it}' x_{it} - \sigma_{it} \leq 0 \text{ for all } h \leq I \text{ and } r \leq t \Big\}. \end{aligned} \tag{6}$$

The value of the objective function at the optimum is an estimate of the reciprocal of the BODF (i.e., an estimate of BOTE). The corresponding dual problem is

$$\begin{aligned} \min_{\mu, \lambda, \tau} \Big\{ \mu :\ & g_{it} \leq \sum_{h=1}^{I} \sum_{r=1}^{t} \lambda_{hr} g_{hr}, \sum_{h=1}^{I} \sum_{r=1}^{t} \lambda_{hr} b_{hr} = \mu b_{it}, \\ & \sum_{h=1}^{I} \sum_{r=1}^{t} (\lambda_{hr} + \tau_{hr}) z_{hr} \leq z_{it},\ \sum_{h=1}^{I} \sum_{r=1}^{t} (\lambda_{hr} + \tau_{hr}) x_{hr} \leq x_{it}, \\ & \sum_{h=1}^{I} \sum_{r=1}^{t} (\lambda_{hr} + \tau_{hr}) = 1, \lambda_{hr} \geq 0 \text{ and } \tau_{hr} \geq 0 \text{ for all } h \text{ and } r \Big\}' \end{aligned} \tag{7}$$

where, again, μ is an unknown scalar and $\lambda = (\lambda_{11}, \ldots, \lambda_{IT})'$ and $\tau = (\tau_{11}, \ldots, \tau_{IT})'$ are unknown vectors. This formulation shows that the DEA estimator scales down the bad-output vector while holding good outputs, inputs, and environmental variables fixed. As before, both the primal and dual problems can be solved using standard DEA software packages by augmenting the dataset with "inactive" variants of the observed firms.

As a simple example, reconsider the industry in which firms use a single input to produce one good output and one bad output while operating in a production environment characterised by a single environmental variable. Fabricated data for six hypothetical firms are reported in Table 5, the same data as in Table 1, reproduced here for convenience. The output combinations in this table can again be depicted as points A to F in Figure 3. In this figure, the DEA frontier is given by the line segments connecting the horizontal axis with frontier points A, B, C and D; this is the same frontier that was shown earlier in Figure 1. DEA estimates of BOTE for each firm are reported in Table 6. Firms A and B have estimated BOTE scores of 1, indicating that they could not have reduced their production of bad outputs while holding inputs and good outputs fixed. By contrast, Firms C and E have BOTE scores of 0.5, indicating that they could have reduced their production of bad outputs by half, Firm D has a BOTE score of 0.095, indicating that it could have reduced its production of bad outputs by $1 - 0.095 = 90.5\%$, and Firm F has a BOTE score of 0.067, indicating that it could have reduced its production of bad outputs by $1 - 0.067 = 93.3\%$. Table 6 also reports estimates of shadow prices and marginal rates of transformation. Again, the MRTs are equal to the ratios of the shadow prices and represent the rate at which the good output can be substituted for the bad output as we move around the frontier. For firms that have BOTE scores less than 1 (e.g., D, E and F), shadow prices and MRTs are computed at projection points on the frontier (e.g., D$''$, E$''$ and F$''$); in the case of the BODF, these projection points are obtained by radially contracting the observed bad-output vector while holding inputs and good outputs fixed. For points A, D and F, the estimated MRT is equal to the slope of the line segment 0A; and for points B, C, and E it is given by the slope of AB. Again, shadow prices—and therefore MRTs—at the corner points (A and B) are not unique.

As another example, consider an industry that uses a single input to produce two bad outputs and one good output in a production environment characterised by a single environmental variable. Fabricated data for six hypothetical firms are reported in Table 7. Because all firms use the same amount of input to produce the same amount of good output, and because they all operate in the same period in the same production environment, their bad output choices can be represented by points A to F in Figure 4. The DEA frontier in this case is the hexagon with vertices at points A to F; this frontier reflects assumptions A1 to A4. Estimates of BOTE for each firm are reported in Table 8. Firms A, E and F have BOTE scores of 1, indicating that they could not have scaled down their production of bad outputs while holding their inputs and good outputs fixed. By contrast, Firms B and D have BOTE scores of 0.667, indicating that they could have reduced their production of bad outputs by one third, while Firm C has a BOTE score of 0.5, indicating that it could have reduced its production of bad outputs by one half. Table 8 also reports shadow prices and marginal rates of transformation. The MRTs are the negatives of the ratios of the shadow prices (i.e., $\text{MRT}_{21} = -\ \text{SPB}_1/\text{SPB}_2$) and represent the decrease in bad output 2 that would accompany a one-unit increase in bad output 1 along the frontier: for points A, B, C and E, the MRT is given by the slope of the line segment AE; and for points D and F it is given by the slope of the line segment EF.

3.3 The Input Distance Function

A period-and-environment-specific IDF gives the reciprocal of the smallest fraction of a given input vector that can be used to produce given good and bad outputs in a given period in a given production environment. For example, if it is possible for a firm to produce its outputs using as little as one half of its inputs, then the IDF takes the value $1/0.5 = 2$. By construction, the IDF takes a value no less than one. Its reciprocal lies in the unit interval and provides a

Table 5: Fabricated Data For Example 3

Firm i	Period t	Input x	Good g	Bad b	Environ. z
A	1	1	3	1	1
B	1	1	5	3	1
C	1	1	5	6	1
D	1	1	2	7	1
E	1	1	4	4	1
F	1	1	1	5	1

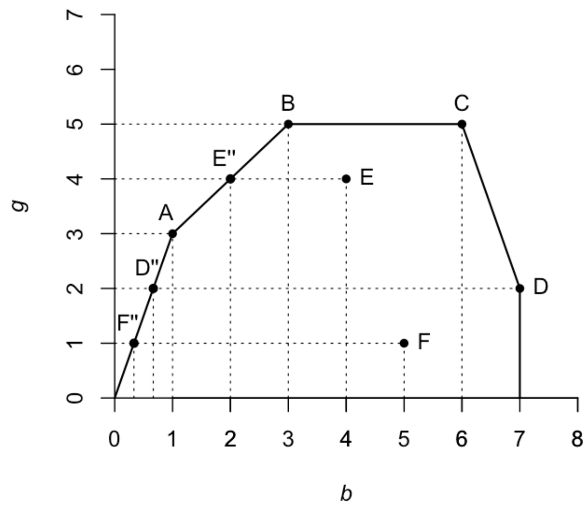

Figure 3: Estimated Output Set For Example 3

Table 6: Economic Quantities of Interest For Example 3

i	BOTE	SPG	SPB	MRT
A	1	0.333	1	3
B	1	0.333	0.333	1
C	0.5	0.167	0.167	1
D	0.095	0.048	0.143	3
E	0.5	0.25	0.25	1
F	0.067	0.067	0.2	3

Table 7: Fabricated Data For Example 4

Firm	Period	Input	Good	Bad 1	Bad 2	Environ.
i	t	x	g	b_1	b_2	z
A	1	1	1	2	4	1
B	1	1	1	3	6	1
C	1	1	1	6	6	1
D	1	1	1	7.5	3	1
E	1	1	1	4	2	1
F	1	1	1	7	2	1

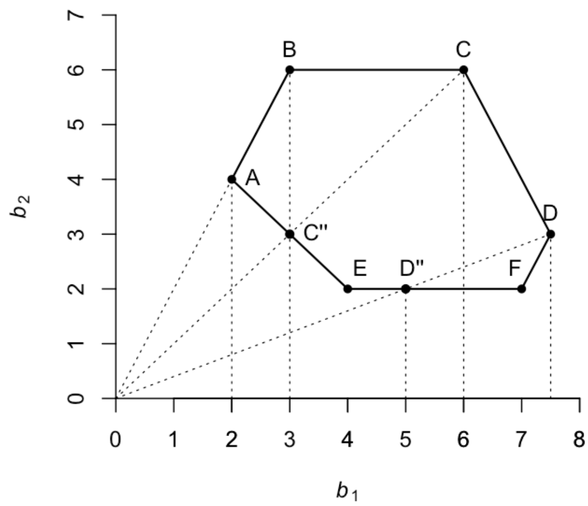

Figure 4: Estimated Output Set For Example 4

Table 8: Economic Quantities of Interest For Example 4

i	BOTE	SPB_1	SPB_2	MRT_{21}
A	1	0.167	0.167	-1
B	0.667	0.111	0.111	-1
C	0.5	0.083	0.083	-1
D	0.667	0	0.333	0
E	1	0.167	0.167	-1
F	1	0	0.5	0

measure of input-oriented technical efficiency (ITE). The ITE of a firm (or its manager) reflects how well inputs are minimised while holding outputs and environmental variables fixed.

If assumptions A1 and A2 are true, then the IDF takes the form:

$$D_I^{it}(x_{it}, g_{it}, b_{it}, z_{it}) = (\theta_{it}' x_{it})/(\xi_{it}' g_{it} - \kappa_{it}' z_{it} - \nu_{it}' b_{it} - \phi_{it}) \tag{8}$$

where ϕ_{it} is an unknown scalar and $\xi_{it} = (\xi_{1it}, \ldots, \xi_{Nit})'$, $\kappa_{it} = (\kappa_{1it}, \ldots, \kappa_{Jit})'$, $\nu_{it} = (\nu_{1it}, \ldots, \nu_{Kit})'$ and $\theta_{it} = (\theta_{1it}, \ldots, \theta_{Mit})'$ are unknown vectors. In this case, DEA estimation involves minimising $D_I^{it}(x_{it}, g_{it}, b_{it}, z_{it})$ subject to constraints that ensure that assumptions A4 to A8 are satisfied. Again, applying the same reasoning as in Section 3.1, the primal LP turns out to be:

$$\begin{aligned}\max_{\xi_{it},\kappa_{it},\phi_{it},\nu_{it},\theta_{it}} \Big\{ & \xi_{it}' g_{it} - \kappa_{it}' z_{it} - \nu_{it}' b_{it} - \phi_{it} : \xi_{it} \geq 0,\ \kappa_{it} \geq 0,\ \theta_{it} \geq 0, \theta_{it}' x_{it} = 1, \\ & \xi_{it}' g_{hr} - \kappa_{it}' z_{hr} - \nu_{it}' b_{it} - \phi_{it} - \theta_{it}' x_{hr} \leq 0 \text{ for all } h \leq I \text{ and } r \leq t, \\ & - \kappa_{it}' z_{hr} - \phi_{it} - \theta_{it}' x_{hr} \leq 0 \text{ for all } h \leq I \text{ and } r \leq t \Big\}.\end{aligned} \tag{9}$$

The value of the objective function at the optimum is an estimate of the reciprocal of the IDF (i.e., an estimate of ITE). The corresponding dual problem is

$$\begin{aligned}\min_{\mu,\lambda,\tau} \Big\{ \mu :\ & g_{it} \leq \sum_{h=1}^{I}\sum_{r=1}^{t} \lambda_{hr} g_{hr}, \sum_{h=1}^{I}\sum_{r=1}^{t} \lambda_{hr} b_{hr} = b_{it}, \\ & \sum_{h=1}^{I}\sum_{r=1}^{t} (\lambda_{hr} + \tau_{hr}) z_{hr} \leq z_{it},\ \sum_{h=1}^{I}\sum_{r=1}^{t} (\lambda_{hr} + \tau_{hr}) x_{hr} \leq \mu x_{it}, \\ & \sum_{h=1}^{I}\sum_{r=1}^{t} (\lambda_{hr} + \tau_{hr}) = 1, \lambda_{hr} \geq 0 \text{ and } \tau_{hr} \geq 0 \text{ for all } h \text{ and } r \Big\}'\end{aligned} \tag{10}$$

where, again, μ is an unknown scalar and $\lambda = (\lambda_{11}, \ldots, \lambda_{IT})'$ and $\tau = (\tau_{11}, \ldots, \tau_{IT})'$ are unknown vectors. This formulation shows that the DEA estimator scales down the input vector while holding outputs and environmental variables fixed. As before, both the primal and dual problems can be solved using standard DEA software packages by augmenting the dataset with "inactive" variants of the observed firms.

As an example, consider an industry in which firms use two inputs to produce one good output and one bad output in a production environment characterised by a single environmental variable. Fabricated data for six hypothetical firms are presented in Table 9. The input choices for these firms correspond to points A to F in Figure 5. Observe that the scatter of points in this figure is identical to that shown earlier in Figure 4; however, the DEA frontier now consists of line segment AE together with the segments extending upward from point A and rightward from point E; this frontier reflects assumptions A1, A2, A4 and A8. Estimates of ITE for each firm are reported in Table 10. Firms A, E and F have ITE scores of 1, indicating that they would not have been able to scale down their inputs while holding their outputs fixed. By contrast, Firms B and D have ITE scores of 0.667, indicating that they could have reduced their inputs by one third, while Firm C has an ITE score of 0.5, indicating that it could have reduced its inputs by one half. Table 10 also reports estimates of the shadow price of input 1

(SPX_1), the the shadow price of input 2 (SPX_2), and the marginal rate of technical substitution ($MRTS_{12} = -\ SPX_1/SPX_2$) for each firm. The MRTSs represent the rate at which input 1 can be substituted for input 2 as we move around the frontier: for points A, B, C and E, the MRTS is given by the slope of the line segment AE; and for points D and F it is given by the slope of the line segment EF.

4 A Related Approach

The estimators introduced in Section 3 are closely related to estimators that are already well established in the DEA literature. To highlight this connection, it is convenient to rewrite problem (4) as

$$\max_{\mu,\lambda,\tau}\{\mu : \ (\mu g_{it}, b_{it}) \in \hat{P}^t(x_{it}, z_{it})\}, \tag{4'}$$

where

$$\begin{aligned}\hat{P}^t(x, z_{it}) = \Big\{(g, b) : \ & g \leq \sum_{h=1}^{I}\sum_{r=1}^{t}\lambda_{hr}g_{hr}, \sum_{h=1}^{I}\sum_{r=1}^{t}\lambda_{hr}b_{hr} = b, \\ & \sum_{h=1}^{I}\sum_{r=1}^{t}(\lambda_{hr}+\tau_{hr})z_{hr} \leq z_{it}, \ \sum_{h=1}^{I}\sum_{r=1}^{t}(\lambda_{hr}+\tau_{hr})x_{hr} \leq x, \\ & \sum_{h=1}^{I}\sum_{r=1}^{t}(\lambda_{hr}+\tau_{hr}) = 1, \lambda_{hr} \geq 0 \text{ and } \tau_{hr} \geq 0 \text{ for all } h \text{ and } r\Big\}\end{aligned} \tag{11}$$

is the DEA estimator of the period-and-environment-specific output set $P^t(x, z_{it})$. This formulation can be simplified in two ways. If the data are cross-section data, then all references to time periods drop out. If there is no environmental change, then the constraints involving the environmental variables can be omitted. In the special case where both conditions hold—i.e., cross-section data with no environmental change—equation (11) collapses to the familiar "activity analysis technology" of Kuosmanen (2005, Eq. 4). Recent empirical applications that implement the Kuosmanen activity-analysis estimator include Shen et al (2022) and Yu and Rakshit (2023).

5 Conclusion

This paper has developed a unified nonparametric framework for measuring firm performance when (i) some variables involved in the production process are outside the control of managers, and (ii) production of goods is accompanied by the production of bads. The proposed DEA estimators yield estimates of frontiers—and associated shadow prices and marginal rates of transformation and technical substitution—that are policy-relevant and free of the biases that arise when environmental variables and bads are ignored. The paper provides primal and dual linear-programming formulations that can be used to estimate three period-and-environment-specific distance functions (good-output, bad-output, and input oriented). Fabricated datasets are used to show how the estimators can be implemented using standard DEA software. Imple-

Table 9: Fabricated Data For Example 4

Firm	Period	Input 1	Input 2	Good	Bad	Environ.
i	t	x_1	x_2	g	b	z
A	1	2	4	1	1	1
B	1	3	6	1	1	1
C	1	6	6	1	1	1
D	1	7.5	3	1	1	1
E	1	4	2	1	1	1
F	1	7	2	1	1	1

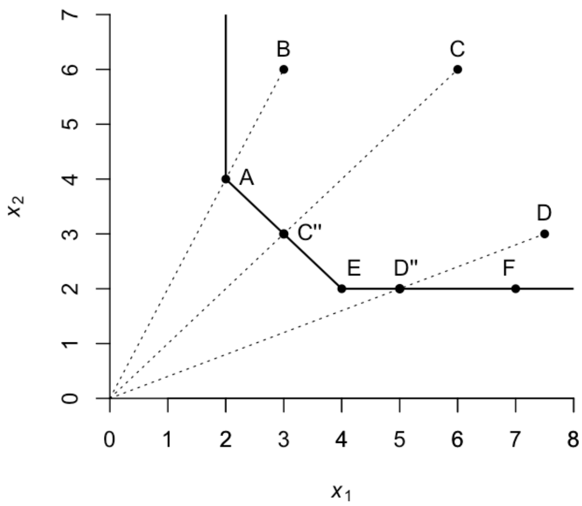

Figure 5: Estimated Output Set For Example 5

Table 10: Economic Quantities of Interest For Example 5

i	ITE	SPX_1	SPX_2	MRTS_{21}
A	1	0.167	0.167	-1
B	0.667	0.111	0.111	-1
C	0.5	0.083	0.083	-1
D	0.667	0	0.333	0
E	1	0.167	0.167	-1
F	1	0	0.5	0

mentation involves augmenting the datasets with inactive variants of the observed firms. The paper makes a connection to an activity-analysis estimator that is already established in the literature, and explains the conditions under which the approaches coincide.

Several limitations suggest promising avenues for further research. First, the DEA estimators developed here assume variable returns to scale and convex production possibilities sets; extensions to constant returns to scale and non-convex settings would broaden the applicability of the approach. Second, alternative disposability assumptions (e.g., costly abatement of bads, by-product linkages, or explicit abatement technologies) could be incorporated to reflect sector-specific physical or regulatory constraints. Finally, given the practical challenges in specifying and measuring inputs, outputs, and environmental variables, there is scope to explore nonparametric estimators that explicitly accommodate noise.

Taken together, the estimators and implementation guidance provided here constitute a practical toolkit for researchers seeking a coherent treatment of environmental variables and bads in frontier estimation. The accompanying code and worked examples are intended to facilitate adoption and adaptation in sectors where these issues are central to both managerial decision-making and policy evaluation.

Appendix: R Code

```
# Table 1 Fabricated Data for Example 1 ----

  Firm = c("A", "B", "C", "D", "E", "F")
  Period = c(1, 1, 1, 1, 1, 1)
  b = c(1, 3, 6, 7, 4, 5)
  g = c(3, 5, 5, 2, 4, 1)
  x = c(1, 1, 1, 1, 1, 1)
  z = c(1, 1, 1, 1, 1, 1)

  df = data.frame(Firm, Period, x, g, b, z)
  colnames(df) = c("i", "t", "x", "g", "b", "z")
  df

# Table 2 Estimatd EQIs for Example 1 ----

  gplus <- rbind(as.matrix(g), 0 * as.matrix(g))
  bplus <- rbind(as.matrix(b), 0 * as.matrix(b))
  xplus <- rbind(as.matrix(x),     as.matrix(x))
  zplus <- rbind(as.matrix(z),     as.matrix(z))

  n    <- nrow(gplus) / 2          # active rows on top
  M    <- ncol(xplus)              # inputs
```

```
  J    <- ncol(zplus)              # env
  K    <- ncol(bplus)              # bads
  N    <- ncol(gplus)              # goods

  qq = cbind(gplus)
  xx = cbind(xplus, zplus, bplus, -bplus)
  pfa = dea(xx, qq, ORIENTATION = "out", DUAL = TRUE)
  GOTE <- 1 / pfa$eff[1:n]
  SPG  <- pfa$vy[1:n, 1:N, drop=FALSE]
  SPB  <- pfa$ux[1:n, (M+J+1):(M+J+K),  drop=FALSE]
        - pfa$ux[1:n, (M+J+K+1):(M+J+2*K), drop=FALSE]
  MRT = SPB / SPG

  df = data.frame(Firm, roundup(cbind(GOTE, SPG, SPB, MRT), 3))
  colnames(df) = c("i", "GOTE", "SPG", "SPB", "MRT")
  df

# Table 3 Fabricated Data for Example 2 ----

  Firm = c("A", "B", "C", "D", "E", "F")
  Period = c(1, 1, 1, 1, 1, 1)
  g1 = c(1, 3, 5, 6, 2, 5)
  g2 = c(3, 6, 4, 2, 2.5, 1)
  x = c(1, 1, 1, 1, 1, 1)
  b = c(1, 1, 1, 1, 1, 1)
  z = c(1, 1, 1, 1, 1, 1)
  g = cbind(g1,g2)

  df = data.frame(Firm, Period, x, g, b, z)
  colnames(df) = c("i", "t", "x", "g1", "g2", "b", "z")
  df

# Table 4 Estimated EQIs for Example 2 ----

  gplus <- rbind(as.matrix(g), 0 * as.matrix(g))
  bplus <- rbind(as.matrix(b), 0 * as.matrix(b))
  xplus <- rbind(as.matrix(x),     as.matrix(x))
  zplus <- rbind(as.matrix(z),     as.matrix(z))

  n    <- nrow(gplus) / 2          # active rows on top
  M    <- ncol(xplus)              # inputs
  J    <- ncol(zplus)              # env
```

```
  K    <- ncol(bplus)              # bads
  N    <- ncol(gplus)              # goods

  qq = cbind(gplus)
  xx = cbind(xplus, zplus, bplus, -bplus)
  pfa = dea(xx, qq, ORIENTATION = "out", DUAL = TRUE)
  GOTE <- 1 / pfa$eff[1:n]
  SPG  <- pfa$vy[1:n, 1:N, drop=FALSE]
  MRT21 = -SPG[,1]/SPG[,2]

  df = data.frame(Firm,roundup(cbind(GOTE, SPG, MRT21),3))
  colnames(df) = c("i", "GOTE", "SPG1", "SPG2", "MRT21")
  df

  sign(SPG)
  # for points D and F, R stores 0 as a negative number, so it
  # computes MRT21 = Inf instead of -Inf.  Could add a command to fix this.

# Table 5 Fabricated Data for Example 3 ----

  Firm = c("A", "B", "C", "D", "E", "F")
  Period = c(1, 1, 1, 1, 1, 1)
  b = c(1, 3, 6, 7, 4, 5)
  g = c(3, 5, 5, 2, 4, 1)
  x = c(1, 1, 1, 1, 1, 1)
  z = c(1, 1, 1, 1, 1, 1)

  df = data.frame(Firm, Period, x, g, b, z)
  colnames(df) = c("i", "t", "x", "g", "b", "z")
  df

# Table 6 Estimated EQIs for Example 3 ----

  gplus <- rbind(as.matrix(g), 0 * as.matrix(g))
  bplus <- rbind(as.matrix(b), 0 * as.matrix(b))
  xplus <- rbind(as.matrix(x),     as.matrix(x))
  zplus <- rbind(as.matrix(z),     as.matrix(z))

  n    <- nrow(gplus) / 2          # active rows on top
  M    <- ncol(xplus)              # inputs
  J    <- ncol(zplus)              # env
  K    <- ncol(bplus)              # bads
```

```
  N     <- ncol(gplus)              # goods

  qq = cbind(gplus,-xplus,-zplus)
  xx = cbind(bplus,-bplus)
  pfa = dea(xx, qq, ORIENTATION = "in",DUAL = TRUE)
  BOTE = pfa$eff[1:n]
  SPG  <- pfa$vy[1:n, 1:N, drop=FALSE]
  SPB  <- pfa$ux[1:n, 1:K, drop=FALSE]
        - pfa$ux[1:n, (K+1):(2*K), drop=FALSE]
  MRT = SPB / SPG

  df = data.frame(Firm, roundup(cbind(BOTE, SPG, SPB, MRT), 3))
  colnames(df) = c("i", "GOTE", "SPG", "SPB", "MRT")
  df

# Table 7 Fabricated Data for Example 4 ----

  Firm = c("A", "B", "C", "D", "E", "F")
  Period = c(1, 1, 1, 1, 1, 1)
  b1 = c(2, 3, 6, 7.5, 4, 7)
  b2 = c(4, 6, 6, 3, 2, 2)
  x = c(1, 1, 1, 1, 1, 1)
  g = c(1, 1, 1, 1, 1, 1)
  z = c(1, 1, 1, 1, 1, 1)
  b = cbind(b1,b2)

  df = data.frame(Firm, Period, x, g, b, z)
  colnames(df) = c("i", "t", "x", "g", "b1", "b2", "z")
  df

# Table 8 Estimated EQIs for Example 4 ----

  gplus <- rbind(as.matrix(g), 0 * as.matrix(g))
  bplus <- rbind(as.matrix(b), 0 * as.matrix(b))
  xplus <- rbind(as.matrix(x),     as.matrix(x))
  zplus <- rbind(as.matrix(z),     as.matrix(z))

  n     <- nrow(gplus) / 2          # active rows on top
  M     <- ncol(xplus)              # inputs
  J     <- ncol(zplus)              # env
  K     <- ncol(bplus)              # bads
  N     <- ncol(gplus)              # goods
```

```
  qq = cbind(gplus,-xplus,-zplus)
  xx = cbind(bplus,-bplus)
  pfa = dea(xx, qq, ORIENTATION = "in",DUAL = TRUE)
  BOTE = pfa$eff[1:n]
  SPB  <- pfa$ux[1:n, 1:K, drop=FALSE]
        - pfa$ux[1:n, (K+1):(2*K), drop=FALSE]
  MRT21 = - SPB[,1]/SPB[,2]

  df = data.frame(Firm,roundup(cbind(BOTE, SPB, MRT21),3))
  colnames(df) = c("i", "BOTE", "SPB1", "SPB2", "MRT21")
  df

# Table 9 Fabricated Data for Example 5 ----

  Firm = c("A", "B", "C", "D", "E", "F")
  Period = c(1, 1, 1, 1, 1, 1)
  x1 = c(2, 3, 6, 7.5, 4, 7)
  x2 = c(4, 6, 6, 3, 2, 2)
  b = c(1, 1, 1, 1, 1, 1)
  g = c(1, 1, 1, 1, 1, 1)
  z = c(1, 1, 1, 1, 1, 1)
  x = cbind(x1,x2)

  df = data.frame(Firm, Period, x, g, b, z)
  colnames(df) = c("i", "t", "x1", "x2", "g", "b", "z")
  df

# Table 10 Estimated EQIs for Example 5 ----

  gplus <- rbind(as.matrix(g), 0 * as.matrix(g))
  bplus <- rbind(as.matrix(b), 0 * as.matrix(b))
  xplus <- rbind(as.matrix(x),     as.matrix(x))
  zplus <- rbind(as.matrix(z),     as.matrix(z))

  n    <- nrow(gplus) / 2          # active rows on top
  M    <- ncol(xplus)              # inputs
  J    <- ncol(zplus)              # env
  K    <- ncol(bplus)              # bads
  N    <- ncol(gplus)              # goods

  qq = cbind(gplus,bplus,-bplus,-zplus)
```

```
xx = cbind(xplus)
pfa = dea(xx, qq, ORIENTATION = "in",DUAL = TRUE)
ITE = pfa$eff[1:n]
SPX = pfa$ux[1:n,]
MRTS21 = -SPX[,1]/SPX[,2]

df = data.frame(Firm,roundup(cbind(ITE,SPX,MRTS21),3))
colnames(df) = c("i", "ITE", "SPX1", "SPX2", "MRTS21")
df
```

References

Ball V, Lovell C, Nehring R, Somwaru A (1994) Incorporating undesirable outputs into models of production: an application to US agriculture. Cahiers d'économie et socoiolgei ruraler 31

Banker R, Morey R (1986) The use of categorical variables in Data Envelopment Analysis. Management Science 32(12):1613–1627

Banker R, Charnes A, Cooper W (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis. Management Science 30(9):1078–1092

Bogetoft P, Otto L (2015) Benchmarking: Benchmark and Frontier Analysis Using DEA and SFA. R package version 2015-7-8

Charnes A, Cooper W, Rhodes E (1978) Measuring the efficiency of decision-making unit. European Journal of Operational Research 2(6):429–444, see also "Corrections" op.cit. 3(4):339

Cooper WW, Ruiz JL, Sirvent I (2007) Choosing weights from alternative optimal solutions of dual multiplier models in DEA. European Journal of Operational Research 180(1):443–458, DOI 10.1016/j.ejor.2006.04.038

Färe R, Grosskopf S, Lovell C, Pasurka C (1989) Multilateral productivity comparisons when some outputs are undesirable: A nonparametric approach. The Review of Economics and Statistics 71(1):90–98

Hailu A, Veeman T (2001) Non-parametric productivity analysis with undesirable outputs: An application to the Canadian pulp and paper industry. American Journal of Agricultural Economics 83(3):605–616

Kuosmanen T (2005) Weak disposability in nonparametric production analysis with undesirable outputs. American Journal of Agricultural Economics 87(4):1077–1082

O'Donnell C (2018) Productivity and Efficiency Analysis: An Economic Approach to Measuring and Explaining Managerial Performance. Springer Nature, Singapore

Seiford L, Zhu J (2002) Modeling undesirable factors in efficiency evaluation. European Journal of Operational Research 142(1):16–20

Shen Z, Balezentis T, Streimikis J (2022) Capacity utilization and energy-related GHG emission in the European agriculture: A data envelopment analysis approach. Journal of Environmental Management 318(115517)

Yu MM, Rakshit I (2023) Target setting for airlines incorporating CO2 emissions: The DEA bargaining approach. Journal of Air Transport Management 108(102376)