Operations Research
Vol. 60, No. 5, September–October 2012, pp. 1064–1079
ISSN 0030-364X (print) | ISSN 1526-5463 (online)

http://dx.doi.org/10.1287/opre.1120.1094

Measuring Eco-Inefficiency: A New Frontier Approach

Chien-Ming Chen
Nanyang Business School, Nanyang Technological University, Singapore 639798, Republic of Singapore, cmchen@ntu.edu.sg

Magali A. Delmas
Institute of the Environment and Sustainability, and Anderson School of Management, University of California, Los Angeles, Los Angeles, California 90095, delmas@ucla.edu

Growing social concerns over the environmental externalities associated with business activities are pushing firms to identify activities that create economic value with less environmental impact and to become more eco-efficient. Over the past two decades, researchers have increasingly used frontier efficiency models to evaluate productive efficiency in the presence of undesirable outputs, such as greenhouse gas emissions or toxic emissions. In this paper, we identify critical flaws in existing frontier models and show that these models can identify eco-inefficient firms as eco-efficient. We develop a new eco-inefficiency frontier model that rectifies these problems. Our model calculates an eco-inefficiency score for each firm and improvements in outputs necessary to attain eco-efficiency. We demonstrate through a Monte Carlo experiment that our eco-inefficiency model provides a more reliable measurement of corporate eco-inefficiency than the existing frontier models. We also extend the single-output Cobb-Douglas production function to multiple desirable and undesirable outputs. This extension allows for greater flexibility in the simulation analysis of frontier models.

Subject classifications: environmental performance; eco-efficiency; nonparametric frontier models; simulation.
Area of review: Environment, Energy, and Sustainability.
History: Received January 2010; revisions received December 2010, October 2011, March 2012; accepted June 2012.

1. Introduction

Increasing social concerns over the environmental externalities of business activities are pushing managers to devise strategies to mitigate their firms' environmental impact (Porter and Reinhardt 2007). Common examples of these strategies include pollution prevention, waste reduction, recycling, closed-loop supply chain management, and environmental management systems (Klassen and McLaughlin 1996, Corbett and Kleindorfer 2001, King and Lenox 2002, Corbett and Klassen 2006, Delmas and Toffel 2008), and managers are faced with the fundamental question of how these mitigating strategies impact corporate performance (King and Lenox 2002, Klassen and Vachon 2003). Because a firm typically utilizes multiple input resources to produce outputs, a variety of input and output variables are required to assess corporate performance assessment. Input variables can include labor, capital assets, investments in new product development, and raw materials. Output variables include products, services, or revenue, as well as undesirable by-products such as greenhouse gas emissions and wastes. The potential trade-off relationships among input and output variables make it very challenging for managers to aggregate these variables and present the information as a simple index, to identify potential improvements, and to facilitate decision-making (Delmas and Doctori-Blass 2010). In this paper, we develop an eco-inefficiency model that aggregates multiple inputs and outputs into an eco-inefficiency score.

Frontier methodologies provide a composite inefficiency score that represents the firm's distance to the best-practice eco-efficiency frontier (Charnes et al. 1978, Shephard 1970). The efficiency frontier includes the firms that produce more desirable outputs with fewer inputs and undesirable outputs than other firms in the sample. The efficiency frontier also indicates the boundary condition that a firm can achieve under the current production technology. Frontier methodologies use a mathematical programming model to extrapolate the efficiency frontier based on the input and output quantities of the sampled firms. A firm's inefficiency score is measured by the improvements in outputs necessary for this firm to reach the extrapolated frontier (i.e., increase desirable output quantities and reduce undesirable output quantities), given the firm's current input level.

Although several studies have developed frontier models to evaluate eco-inefficiency (e.g., Hailu and Veeman 2001, Färe et al. 1989, Chung et al. 1997, Seiford and Zhu 2002), our analysis shows that the current frontier models may have significant flaws. Specifically, under the current frontier models, inefficient firms may be identified as eco-efficient, firms' inefficiency scores may improve with an increase in undesirable outputs (i.e., nonmonotonic in undesirable outputs), and inefficiency scores are insensitive to changes in undesirable outputs. These models, however, have been widely used and a bibliographical search in Google Scholar shows that these existing frontier models have been cited more than 1,200 times.[1]

In this paper we build on the nonparametric frontier approach to develop an eco-inefficiency model that overcomes the validity problems of current frontier models. We use a Monte Carlo simulation experiment to compare the performance of our eco-inefficiency model and four representative frontier models as recently identified by Hua and Bian (2007), namely, the directional distance function (DDF model; Chung et al. 1997), the hyperbolic model (Färe et al. 1989), the Seiford and Zhu model (SZ model; Seiford and Zhu 2002), and the "undesirable output as input" (UINP) model (Hailu and Veeman 2001). The simulation results show that our eco-inefficiency model outperforms current representative frontier models across different parameters settings. We also show that our eco-inefficiency model produces a more precise assessment of the inefficiency effect than other existing models.

This new model can be used to analyze undesirable outputs beyond the environmental context. Undesirable outputs are a consequence of many corporate operations. Undesirable outputs include debts or loans, accidents, delays, corporate social irresponsibility, defective products, and waste; see Chen and Delmas (2011) and Chen et al. (2010) for a more comprehensive discussion and references. Our model can prove useful for evaluating operational efficiencies in these contexts. Our paper also extends the frontier model simulation methodology, which has been limited to a single desirable output, to include multiple desirable and undesirable outputs. Our simulation model allows for greater flexibility in the analysis of frontier models for measuring eco-inefficiency by integrating environmental and productive inefficiencies.

In the next section we introduce the general frontier methodology and the four frontier models that have been developed to handle undesirable outputs. In §3 we present our eco-inefficiency model and demonstrate its advantages. In §4, we use a Monte Carlo simulation to compare the performance of our eco-inefficiency model with the other frontier models. In §5 we summarize our findings and contributions.

2. Frontier Methodology and Existing Models

2.1. Fundamental Concepts of Frontier Models

The nonparametric frontier model, also known as data envelopment analysis (DEA), uses linear programming to aggregate multiple inputs and outputs of firms into a relative efficiency score (Charnes et al. 1978, Cooper et al. 2007). The set of feasible production plans, or technology set, are the input–output combinations enveloped by the frontier. If a firm is on this frontier, it is considered efficient (Shephard 1970, Banker et al. 1984). If a firm is not on the frontier, the distance to the best-practice frontier represents the firm's inefficiency.

We now describe the efficient frontier model in a linear programming form. In the model, we consider three vectors. The inputs $X = (x_1, \ldots, x_M)$ are the resources used to produce the desirable outputs $Y = (y_1, \ldots, y_N)$ and undesirable outputs $U = (u_1, \ldots, u_P)$. Given that we observe K firms in our sample, the production technology set can be formulated as follows (Shephard 1970, Charnes et al. 1978, Färe and Grosskopf 2004):

$$\Omega = \{(X, Y, U)\colon X \text{ can produce } Y \text{ and } U\} \tag{1.1}$$

$$= \Bigg\{(X, Y, U)\colon \sum_{k=1}^{K} z_k x_{km} \leqslant x_m \quad \text{for } m = 1, \ldots, M, \tag{1.2}$$

$$\sum_{k=1}^{K} z_k y_{kn} \geqslant y_n \quad \text{for } n = 1, \ldots, N, \tag{1.3}$$

$$\sum_{k=1}^{K} z_k u_{kp} = u_p \quad \text{for } p = 1, \ldots, P, \tag{1.4}$$

$$z_k \geqslant 0 \quad \text{for } k = 1, \ldots, K\Bigg\}, \tag{1.5}$$

where $(x_{k1}, \ldots, x_{kM})$, $(y_{k1}, \ldots, y_{kN})$, and $(u_{k1}, \ldots, u_{kP})$ are the input and output vectors of the kth firm in the sample, and z_k is the intensity variable associated with the kth firm. The z_k variable indicates the importance of the kth firm in constructing the efficient frontier for a specific point (X, Y, U) in the production set.

The constraints (1.2) to (1.5) form a polyhedron also referred to as the production set, which is the collection of feasible inputs and outputs (X, Y, U) under the current production technology. Points in the production set are those achievable under the current technology constraints. The production set as defined in (1) has the following properties (Färe and Grosskopf 2004):

Property 1. $(X, Y, U) \in \Omega$ *and* $X_1 \geqslant X$ *imply* $(X_1, Y, U) \in \Omega$.

Property 2. $(X, Y, U) \in \Omega$ *and* $Y_1 \leqslant Y$ *imply* $(X, Y_1, U) \in \Omega$.

Property 3. $(X, Y, U) \in \Omega$ *implies* $(\theta X, \theta Y, U) \in \Omega$ *for* $0 \leqslant \theta \leqslant 1$.

A key presumption underlying these three properties is that, if (X, Y, U) is observed, this observation is by definition a member of the production set (i.e., the axiom of "inclusion of observations"). With these three properties, the frontier methodologies extrapolate the entire production set based on the input–output observations in the sample. The first two properties mean that, if (X, Y, U) is observed, then using more inputs to produce a smaller amount of desirable outputs (i.e., (X_1, Y_1, U)) is also feasible. This is called the *strong disposability assumption*, because inputs or outputs can change unilaterally without compromising each other. If undesirable outputs can be generated without subsequent cost or damage, undesirable outputs are said to be strongly disposable and the production set is the same as (1) except we replace (1.5) with (2):

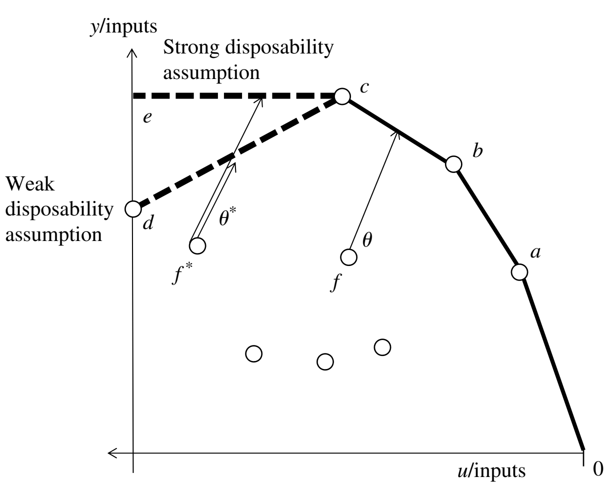
$$\sum_{k=1}^{K} z_k u_{kp} \leqslant u_p \quad \text{for } p = 1, \ldots, P. \tag{2}$$

By contrast, under the *weak disposability assumption*, a reduction in undesirable outputs of (X, Y, U) will result in a reduction of desirable outputs. This property is expressed through the equality constraint (1.4). Here, weak disposability only applies to undesirable outputs, because we assume that producers cannot dispose freely of the undesirable outputs. For example, electric utility plants may need to invest in carbon capture devices to reduce their greenhouse gas emissions (Gibbins and Chalmers 2008).

Figure 1 displays a production set with a desirable output y and an undesirable output u to illustrate how to compute an inefficiency score. The horizontal axis represents the undesirable output u and the vertical axis represents the desirable output y. We divide the output quantity of each firm by its input quantity to evaluate firms' eco-inefficiency based on y and u. Firms with a high eco-efficiency are those situated in the upper-right corner of the graph, where they produce more desirable outputs and low undesirable outputs. We use piecewise-linear segments to extrapolate the eco-efficient frontier by linking firms in the upper-right corner. Firms on the frontier are considered eco-efficient because no other firms in the production set can produce more desirable outputs and fewer undesirable outputs.

In Figure 1, the frontier is the line segment "0*abcd*" if we assume that the undesirable output is weakly disposable, and is "0*abce*" if the undesirable output is strongly disposable. It is also important to note that the *cd* portion of the efficient frontier is dominated by the point c; i.e., c produces a higher quantity of the desirable output y and a lower quantity of the undesirable output u. We call the *cd* portion of the frontier the *misspecified efficient frontier*. However, the "*ce*" portion of the frontier (under the strong disposability assumption) is not fully efficient, because points on "*ce*" produce as many desirable outputs as c does, but they produce more undesirable outputs than c. As we explained earlier, in the strong disposability assumption, undesirable outputs are free, and therefore firms do not need to allocate resources to compensate for the emissions of undesirable outputs. The difference in disposability assumptions is characterized by the inequality signs for the undesirable output constraints (1.4). As a result, the production set associated with the strong disposability assumption is larger than the production set under the weak disposability assumption (see Figure 1).

Figure 1. Illustration of the frontier model under different disposability assumptions.



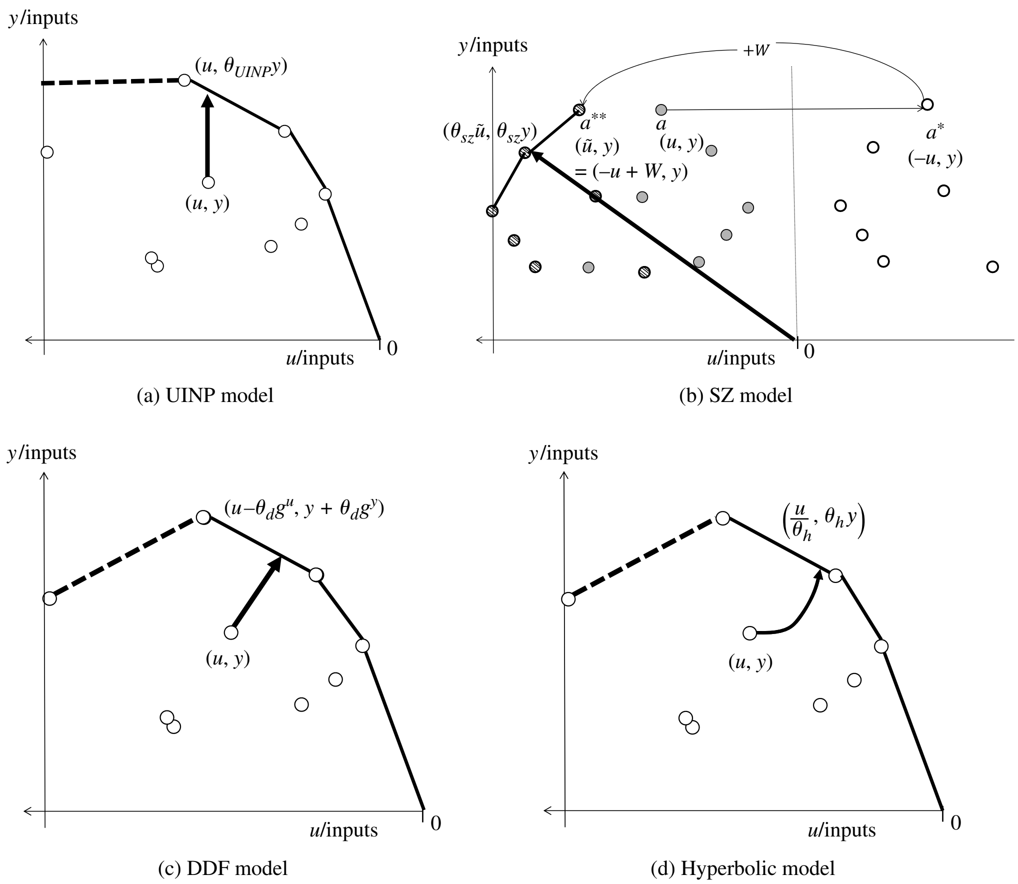
In a frontier model, the inefficiency score of a firm represents the firm's distance to the efficient frontier. The inefficiency score is computed as an optimization problem, in which the efficient frontier is the boundary of the feasible region. An objective function determines both the direction of the evaluated firm toward the frontier, as well as how the distance between the firm and the frontier is calculated. As shown in Figure 1, different frontier models may adopt different assumptions (e.g., weak disposability or strong disposability assumption) and different objective functions to calculate inefficiency scores. The inefficiency score and the benchmark target depend on these two settings. For example, firm f in Figure 1 is eco-inefficient because it is not on the efficient frontier, whereas firms a, b, and c are eco-efficient. However, firm f can move in different directions to reach the frontier. Next we introduce four frontier models for undesirable outputs.

2.2. Current Frontier Models for Undesirable Outputs

In this section we give a brief overview of four representative frontier models that incorporate undesirable outputs: the directional distance function (DDF model; Chung et al. 1997), the hyperbolic model (Färe et al. 1989), the Seiford and Zhu model (SZ model; Seiford and Zhu 2002), and the "undesirable output as input" (UINP) model (Hailu and Veeman 2001). We first introduce the UINP and SZ models, which both assume strong disposability on undesirable outputs and use the traditional DEA model mathematical formulation (Charnes et al. 1978). Second we introduce the DDF and hyperbolic models, which assume weak disposability on undesirable outputs. This second set of models has been used widely in various industry contexts, including banks, electricity industries, industry efficiency, provincial governments, agriculture, and airports; see Chen et al. (2010) for a discussion and references to these applications. Figure 2 provides graphic illustrations of these four models.

The UINP model treats undesirable outputs as inputs, because firms are expected to minimize their input

Figure 2. Illustration of the four frontier models for undesirable outputs.



consumption (Hailu and Veeman 2001). The UINP model is therefore identical to the traditional DEA model; namely,

$$\max\{\theta_{\mathrm{UINP}} \mid (X, \theta_{\mathrm{UINP}} Y, U) \in \Omega_{\mathrm{UINP}}\} \tag{3}$$

where Ω_{UINP} is constructed by replacing the (1.4) of Ω with $\sum_{k=1}^{K} z_k u_{kp} \leqslant u_p$ for $p = 1, \ldots, P$; i.e., the strong disposability condition. The inefficiency score (i.e., the optimal value of (3), denoted by θ^*_{UINP}) represents the extent to which a firm can scale up its desirable outputs, given its current inputs and undesirable outputs. For this reason, the UINP model has been criticized for not accurately representing the production process, because undesirable outputs are modeled as inputs (Seiford and Zhu 2002). In Figure 2(a), we observe that the UINP inefficiency score measures the shortfall of desirable outputs Y, given a firm's current level of inputs and undesirable outputs. The UINP model assumes that firms should improve their eco-efficiency by increasing desirable outputs, but not by reducing undesirable outputs.

Seiford and Zhu (2002) take a more heuristic approach to undesirable outputs. The SZ model substitutes undesirable output variables by auxiliary output variables. These new variables are computed by adding a positive scalar to the original undesirable outputs after multiplying them by -1. The SZ model transforms undesirable output variables as:

$$\tilde{U} = -U + W, \tag{4}$$

where W is a predetermined vector making the new undesirable vector $\tilde{U}$ positive for all firms. As shown in Figure 2(b), first the undesirable output vector U is multiplied by -1 (e.g., $a^* = -a$ in Figure 2(b)). A translation vector W is then added to the negative vector $-U$ so that $\tilde{U}$ is strictly positive (e.g., $a^{**} = a^* + W$ in Figure 2(b)). The new undesirable vector $(X, Y, \tilde{U})$ used to construct the production set $\tilde{\Omega}$ under the strong disposability assumption (e.g., (2)). Thus, maximizing these transformed output variables is equivalent to reducing the undesirable outputs. The inefficiency score of the SZ model is obtained from (5)

$$\max\{\theta_{\mathrm{SZ}} \mid (X, \theta_{\mathrm{SZ}} Y, \theta_{\mathrm{SZ}} \tilde{U}) \in \tilde{\Omega}\}. \tag{5}$$

By maximizing the objective function θ_{SZ} in (5), we are scaling up Y and scaling down U at the same time (see Figure 2(b)). However, the inefficiency score (i.e., the optimal value θ^*_{SZ} of (5)) may depend on the choice of translation vector W (Sahoo et al. 2011).

Table 1. Modeling assumptions and ranges of scores.

Models	Assumptions on undesirable outputs	Score range	Limitations
Hyperbolic efficiency	Weakly disposable model	$[1, \infty)$	The model is a nonlinear optimization problem, and therefore if the problem size is large it could be difficult to solve.
Directional distance function (DDF)	Weakly disposable	$[0, \infty)$	The model requires us to specify a directional vector beforehand, and the inefficiency score varies for different choices of directional vectors.
Undesirable output as input model (UINP)	Treated as inputs	$[1, \infty)$	Not representative of the production process; the model cannot provide the benchmark value for undesirable outputs.
Seiford and Zhu's model (SZ)	Strongly disposable; undesirable outputs are transformed before computation.	$[1, \infty)$	The model requires us to specify a translation vector beforehand, and the inefficiency score varies for different choices of translation vectors.

Unlike the UINP and SZ models, the DDF and hyperbolic models impose a weak disposability assumption on undesirable outputs. They have identical production sets, but differ in their inefficiency indexes, which are illustrated in Figures 2(c) and 2(d). In the DDF model, firms follow a predetermined direction (g^Y, g^U) towards the frontier; the inefficiency score θ_d^* is the optimal value of problem (6):

$$\max\left\{\theta_d \mid (X, Y+\theta_d g^Y, U-\theta_d g^U) \in \Omega\right\}. \tag{6}$$

In the DDF model, the inefficiency scores may vary with different directional vectors (Färe and Grosskopf 2004). In the hyperbolic model, the inefficiency is measured by expanding the firm's desirable outputs and contracting undesirable outputs by the same factor. The inefficiency score θ_h^* is the optimal value of (7):

$$\max\{\theta_h \mid (X, Y\theta_h, U/\theta_h) \in \Omega\}. \tag{7}$$

The locus of projecting a firm to the efficient frontier is hyperbolical; see Figure 2(d). Note that the hyperbolic model is a nonlinear and nonconvex optimization problem, and therefore the model is difficult to solve, especially for a large sample.

The modeling assumptions and ranges of efficiency scores of these four models are summarized in Table 1. Note that in all four models, the efficiency status is achieved when a firm obtains the lower-bound value (i.e., 1 or 0), which means that further expansion of desirable outputs and reduction of undesirable outputs is impossible.

An electronic companion to this paper is available as part of the online version at http://dx.doi.org/10.1287/opre.1120.1094. The online Appendix B in the electronic companion contains an illustrative application of the four existing models, in which we use data from paper mill production to test these four models. We find that these models not only fail to capture actual fluctuations in undesirable outputs, but also tend to produce misleading efficiency measurement results. Importantly, the results show that both hyperbolic and DDF models are *not* monotonic in undesirable outputs (i.e., increasing pollution quantities can improve eco-efficiency scores and vice versa), which is contrary to the general beliefs in production economics such as those stated in Färe et al. (2005).

3. Mathematical Formulations

3.1. Eco-Inefficiency Model

In this section we show that in the DDF and hyperbolic models firms' eco-efficiency may improve with an increase in undesirable outputs. We then propose a model that corrects this issue.

We first show graphically the intuition behind our model, and then present the mathematical formulations of our model. In Figure 1, firm f obtains an inefficiency score of θ under the weak disposability assumption. When we increase firm f's undesirable outputs to f^*, the inefficiency score becomes θ^*, which is closer to the efficiency frontier under a weak disposability assumption. Under a strong disposability assumption the efficient frontier is "0*abce*," whereas under the weak disposability assumption it is "0*abcd*." Clearly, θ is larger than θ^*, and hence, under a weak disposability assumption, firm f^* appears to be more efficient than f. If firm f increases its undesirable output further, it can overtake firm d and becomes efficient.

The reverse situation is similarly problematic: if a firm manages to cut its undesirable output from the position of f^* to f, it is considered less efficient in the model. We can attribute this issue to the characteristics of the predetermined directional vector or hyperbolic curve of the conventional efficiency measure. Our model overcomes this issue by allowing firms to select their own directions for improvement to reach the efficiency frontier.

Our eco-inefficiency model is presented below (for each observation "o"):

$$E(x_{om}, y_{on}, u_{op}) = \max \frac{1}{N+P}\left\{\sum_{n=1}^{N}\frac{g_n^y}{y_{on}} + \sum_{p=1}^{P}\frac{g_p^u}{u_{op}}\right\}, \tag{8.1}$$

$$\sum_{k=1}^{K} z_k x_{km} \leqslant x_{om} \quad \text{for } m = 1, \ldots, M, \tag{8.2}$$

$$\sum_{k=1}^{K} z_k y_{kn} \geqslant y_{on} + g_n^y \quad \text{for } n = 1, \ldots, N, \tag{8.3}$$

$$\sum_{k=1}^{K} z_k u_{kp} \leqslant u_{op} - g_p^u \quad \text{for } p = 1, \ldots, P, \tag{8.4}$$

$$z_k \geqslant 0, \quad g_n^y \geqslant 0, \quad g_p^u \geqslant 0 \quad \text{for all } k, n, p. \tag{8.5}$$

The eco-inefficiency model uses an additive inefficiency index similar to the DDF model (i.e., (8.3) and (8.4)). This additive inefficiency index can be contrasted with the radial inefficiency index in the UINP and SZ models, which assume that the evaluated firm could reach the efficiency frontier by proportionally changing its undesirable and desirable outputs. However, in practice there is no guarantee that firms can improve their efficiency by decreasing undesirable outputs and increasing desirable outputs proportionally. Thus, this assumption may be unrealistic in many situations. Another advantage of model (8) is that the benchmark target derived from the slacks-based model such as (8) must be efficient regardless of the type of disposability assumption, whereas the radial inefficiency measure can identify dominated points as benchmark targets (Cooper et al. 1999, Tone 2001, Chen 2012). We maximize the objective function in order to assure that the evaluated firm is benchmarked with an efficient frontier point. The variables g_n^y and g_p^u in model (8.1) represent the amount of output improvements that the evaluated firm can make to reach its benchmark target on the efficiency frontier. Correspondingly, the objective function is the average magnitude of these improvements. For example, a score of 0.5 means that the firm can increase its desirable outputs by 50% and reduce undesirable outputs by 50%.

The objective value of Equation (8.1) represents the overall degree of output inefficiency. It is calculated as the average amount of potential output improvement divided by the observed output value, y_{on} and u_{op} in (8.1). The inefficiency score in theory can take value from zero to infinity. However, if $g_n^y/y_{on} \leqslant 1$ for all n, which is usually true in practice, the score then has an upper bound of 1. A score of zero value means that the evaluated firm is on the efficiency frontier and has no output slacks (hence the firm is *efficient*). If a firm's score is positive, the larger the value, the more inefficient the firm is. Model (8) is also a linear approximation of the classical nonoriented Russel measure (Cooper et al. 2007, pp. 102–104), in which the evaluated firm's inputs are contracted and outputs are expanded much like in the hyperbolic model, but in Russel measure each input and each output are allowed to choose a different expansion or contraction factor (Cooper et al. 1999).

We illustrate our model in Figure 3, where we consider one desirable and one undesirable output. In our model, instead of using a fixed direction to reach the frontier, the evaluated firm (u, y) is free to choose an improvement direction that maximizes its potential for improvement, and therefore its efficiency. This flexibility is consistent with a fundamental concept of the efficiency frontier: because every point on the efficient frontier is considered *efficient*, different production mixes, as represented by different points on the frontier, should appear equally "attractive" for inefficient firms. We show in the next section that the flexibility to choose an improvement direction helps avoid issues associated with both a weak or strong disposability assumption on undesirable outputs (i.e., result in output-dominated benchmark targets).

The eco-inefficiency score provides an aggregate measure of a firm's relative efficiency compared to other firms in the sample. After solving the eco-inefficiency model, we can also identify the efficiency target that the evaluated firm can emulate. Specifically, the benchmark target for firm o can be obtained as:

$$(x_{om}, y_{on} + g_n^{y*}, u_{op} - g_p^{u*}) \quad \text{for all } m, n \text{ and } p, \tag{9}$$

where (g_n^{y*}, g_p^{u*}) is the optimal solution to model (8).

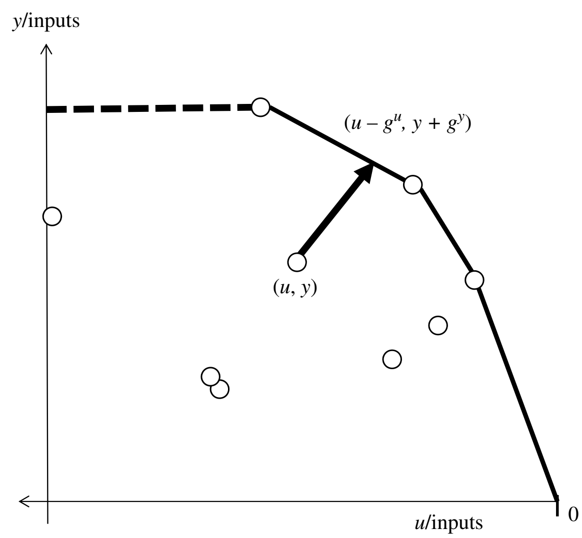
3.2. Properties of the Eco-Inefficiency Model

In this section we show some important properties of the model. Proofs of these results are provided in the online Appendix C, available at http://dx.doi.org/10.1287/opre.1120.1094. Theorem 1 shows that our eco-inefficiency model is unit invariant in inputs and outputs:

THEOREM 1. $\mathbf{E}(\mathbf{x_{om}}, \mathbf{y_{on}}, \mathbf{u_{op}})$ *is homogeneous of degree zero in* $\mathbf{x_{om}}\mathbf{y_{on}}$, *and* $\mathbf{u_{op}}$; *i.e., if we replace the original data* $(\mathbf{x_{om}}, \mathbf{y_{on}}, \mathbf{u_{op}})$ *by* $(\boldsymbol{\alpha}\mathbf{x_{om}}, \boldsymbol{\beta}\mathbf{y_{on}}, \boldsymbol{\gamma}\mathbf{u_{op}})$ *for all k, where* $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, *and* $\boldsymbol{\gamma}$ *are positive numbers, we still have* $\mathbf{E}(\mathbf{x_{om}}, \mathbf{y_{on}}\mathbf{u_{op}}) = \mathbf{E}(\boldsymbol{\alpha}\mathbf{x_{om}}, \boldsymbol{\beta}\mathbf{y_{on}}, \boldsymbol{\gamma}\mathbf{u_{op}})$ *for each observations* $\mathbf{o}$.

The homogeneity (or unit invariance) property is useful because it facilitates comparisons of efficiency across different measurement systems. The "unit-less" property of efficiency scores has also long been recognized as important in engineering and science; see the discussion and examples in Cooper et al. (2007, Chapter 1). Without the homogeneous property, the inefficiency scores would depend on the unit of measurement (e.g., in pounds, kg, or tons; or in Euros or dollars). This would make the interpretation and comparison of the scores more difficult. Traditional DEA models, where all outputs are desirable outputs,

Figure 3. Illustration of the proposed eco-inefficiency model.



have the homogeneous property (Charnes et al. 2007). We can easily verify that the DDF, hyperbolic, SZ, and the UINP models also possess the homogeneous property.

Another important property that needs to be carefully verified is the quality of the eco-inefficiency measure. Ideally, we would expect that eco-efficient firms, as identified by the model, should be "at least as good as" any members in the technology set. Conversely, firms will be considered inefficient only when they are dominated by at least one point in the technology set. To answer this question, we need to first define the dominance relationship in the technology set.

Definition 1. The production plan $(\mathbf{x}_{om}, \mathbf{y}_{on}, \mathbf{u}_{op}) \in \boldsymbol{\Omega}$ is nondominated in outputs if there does not exist any $(\mathbf{x}_{om}, \mathbf{y}'_{on}, \mathbf{u}'_{op}) \in \boldsymbol{\Omega}$ such that $(\mathbf{x}_{om}, \mathbf{y}'_{on}, \mathbf{u}'_{op}) \neq (\mathbf{x}_{om}, \mathbf{y}_{on}, \mathbf{u}_{op}) \in \boldsymbol{\Omega}$ while $\mathbf{y}'_{on} \geqslant \mathbf{y}_{on}$ and $\mathbf{u}'_{on} \leqslant \mathbf{u}_{on}$. Otherwise, $(\mathbf{x}_{om}, \mathbf{y}_{on}, \mathbf{u}_{op}) \in \boldsymbol{\Omega}$ is dominated.

The next theorem shows that the eco-efficiency status is equivalent to the nondominance status in the technology set.

Theorem 2. $\mathbf{E}(\mathbf{x}_{om}, \mathbf{y}_{on}, \mathbf{u}_{op}) = \mathbf{0}$ *if and only if* $(\mathbf{x}_{om}, \mathbf{y}_{on}, \mathbf{u}_{op}) \in \boldsymbol{\Omega}$ *is nondominated in outputs.*

Theorem 2 implies that our eco-inefficiency model always identifies nondominated benchmark target points. Graphically, it means that the eco-inefficiency model always locate points on the efficiency frontier as benchmark target points (see Figure 3). Algebraically, Theorem 2 implies that the constraints on undesirable outputs (8.4) are always binding, and therefore eco-inefficiency scores from model (8) do not depend on the type of disposability assumptions imposed on undesirable outputs. This characteristic separates our eco-inefficiency model from other competing models described earlier, in that our model does not depend on the disposability assumption, and by Theorem 2 we can show that the benchmark targets (i.e., Equation (9)) from our eco-inefficiency model are always nondominated. By contrast, the benchmark targets for some firms under the weak disposability assumption may be the following: given the input vector X, the DDF and hyperbolic models may result in a benchmark point (u^*, y^*) satisfying $\sum_{k=1}^{K} z_k^* y_{kn} > y_n^*$ for $n = 1, \ldots, N$, and $\sum_{k=1}^{K} z_k^* u_{kp} = u_p^*$ for $p = 1, \ldots, P$, where z_k^*'s are the optimal solution from these two models and (u^*, y^*) is calculated according to model (6) or (7). The (u^*, y^*) in this situation is dominated in desirable outputs by the efficiency frontier (i.e., the left-hand side values of the constraints). Formally, the equality constraints limit the solution space of z_k and how the other observations (i.e., all the inputs and outputs on the left-hand side of constraints) can span the benchmark point (u^*, y^*). As noted earlier (see also Figure 1), it is for the same reason that the benchmark point may even be dominated in terms of desirable and undesirable outputs.

For instance, now consider a simple example, where we evaluate an observation $(x, y_1, y_2, u_1, u_2) = (1, 5, 5, 15, 10)$ against two other observations $(1, 10, 11, 5, 7)$ and $(1, 11, 10, 7, 5)$. The first observation is dominated by the other two in all desirable and undesirable outputs, but the first observation is considered environmentally efficient in both the hyperbolic and DDF models (with all components of the directional vector set to 1; i.e., $g^y = g^u = 1$). This also suggests that the first observation is its own benchmark point. On the other hand, when the strong disposability assumption is imposed, the benchmark targets (u^*, y^*) may be weakly dominated in outputs (Cooper et al. 2007; see Figure 1). This limitation holds for the UINP and the SZ models, because they both assume a strong disposability on undesirable outputs.

Theorem 2 allows us to check whether a firm has been misclassified as an efficient firm in the DDF and hyperbolic models:

Corollary 1. *If firm* **"o"** *is efficient in the DDF or hyperbolic model* ($\boldsymbol{\theta}_{\mathbf{d}}^* = \mathbf{0}$ *or* $\boldsymbol{\theta}_{\mathbf{h}}^* = \mathbf{1}$) *but inefficient in the eco-inefficiency model* (*i.e.,* $\mathbf{E}(\mathbf{x}_{om}, \mathbf{y}_{on}, \mathbf{u}_{op}) > \mathbf{0}$), *then firm "***o***" is dominated in outputs in* $\boldsymbol{\Omega}$.

Corollary 1 applies to firms located on the misspecified efficient frontier due to the weak disposability assumption (see Figure 1 the "*cd*" line). These firms are dominated points in the production set, but in the DDF and hyperbolic models these firms may be identified as efficient (see Figures 1 and 2). If a firm appears efficient in these two models but inefficient in the eco-inefficiency model, this firm must be dominated (therefore inefficient) in the production set. In the application to the paper mill production data in the online Appendix B, firms whose efficient targets are on the misspecified efficient frontier in the DDF and hyperbolic models can obtain distorted inefficiency scores (see Figure 1). We can similarly verify whether a firm has the above problem by calculating their efficient targets under these two models. Then we can apply Corollary 1 and verify whether the firm's eco-inefficiency score is equal to 0.

4. Monte Carlo Experiment

The paper mill data analysis presented in the online appendix offers some initial evidence about the drawbacks of the current frontier approaches for eco-efficiency. To further explore these limitations, we employ a Monte Carlo experiment to compare our model with the other four frontier models. To be comprehensive, we also include a hybrid approach that combines our eco-inefficiency model (8) and the SZ model (5). Specifically, the hybrid model presented in (10) integrates the technique of translating the undesirable outputs as were used in the SZ model (5) and the additive efficiency measure as in our eco-inefficiency model (for each observation "o"):

$$E(x_{om}, y_{on}, u_{op}) = \max \frac{1}{N+P} \left\{ \sum_{n=1}^{N} \frac{g_n^y}{y_{on}} + \sum_{p=1}^{P} \frac{g_p^u}{u_{op}} \right\},$$

$$\sum_{k=1}^{K} z_k x_{km} \leqslant x_{om} \quad \text{for } m = 1, \ldots, M,$$

$$\sum_{k=1}^{K} z_k y_{kn} \geqslant y_{on} + g_n^y \quad \text{for } n = 1, \ldots, N,$$

$$\sum_{k=1}^{K} z_k \tilde{u}_{kp} \geqslant \tilde{u}_{kp} + g_p^u \quad \text{for } p = 1, \ldots, P,$$

$$z_k \geqslant 0, \quad g_n^y \geqslant 0, \quad g_p^u \geqslant 0 \quad \text{for all } k, n, p, \tag{10}$$

where $\tilde{u}_{kp} = -u_{kp} + W_p$ for all k and W_p is a parameter satisfying $W_p > \max_k\{u_{kp}\}$ for all p. Like the SZ model (5), the hybrid model (10) translates the undesirable outputs into $\tilde{u}_{kp}$. We next describe the production function used in the simulation.

4.1. Production Function

In the production economics literature, researchers have typically utilized the Cobb–Douglas production function to generate the input and output samples because of its flexibility and simplicity (e.g., Golany and Tamir 1995, Zhang and Bartels 1998, Bardhan et al. 1998, Coelli et al. 2005, Banker and Natarajan 2008, Kuosmanen and Johnson 2010). Specifically, we use the two-input Cobb-Douglas model:

$$\begin{aligned} \log y &= f(x_1, x_2) + \nu - \mu \\ &= \log \alpha_0 + \alpha_1 \log x_1 + \alpha_2 \log x_2 + \nu - \mu. \end{aligned} \tag{11}$$

In Equation (11), the output quantity (y) is the sum of the Cobb-Douglas function $f(x_1, x_2)$, a random noise term (ν), less the inefficiencies (μ) in the production process; x_1 and x_2 represent two distinct inputs, whereas α_1, α_2 are the parameters of the production function. This function corresponds to the maximal output quantity that is technically achievable by using (x_1, x_2). The Cobb-Douglas function exhibits increasing returns-to-scale (RTS) if $\alpha_1 + \alpha_2 > 1$, constant RTS if $\alpha_1 + \alpha_2 = 1$, and decreasing RTS if $\alpha_1 + \alpha_2 < 1$ (Coelli et al. 2005). Then the function $f(x_1, x_2)$ forms the efficient frontier that we use to benchmark firm performance. The term ν stands for sampling errors as commonly seen in most econometric models, and μ represents the inefficiency effect. The random variable ν is typically assumed to follow a standard normal distribution, whereas μ is assumed to follow a one-sided distribution, such as a half-normal distribution, and is nonnegative (Coelli et al. 2005).

We illustrate the production function in Figure A.1 in the online Appendix A. In Figure A.1 we plot a hypothetical Cobb–Douglas production function with one input x and one output y. Observed input–output quantities are represented by asterisks located on the upper and lower side of the production frontier. Figure 4 also represents production functions with increasing, constant, and decreasing returns-to-scale. The deviation from the production function (e.g., $y^* - y_0$) results from the joint influence of the noise and inefficiency terms (i.e., $\exp(\nu - \mu)$).

The production function (11) leads to a single output. However, the evaluation of eco-efficiency requires the consideration of multiple outputs and a model that can integrate both desirable and undesirable outputs. One approach used in prior studies is to model undesirable outputs as inputs in the production function (Koop 1998). This approach is akin to the UINP model and therefore is endowed with similar limitations (see Table 1). To avoid these potential limitations, Fernández et al. (2002) use two production functions to estimate the technical and environmental efficiencies separately (i.e., the production of desirable and undesirable outputs, respectively). The production function of desirable outputs depends on inputs only and the production function of undesirable outputs depends on desirable outputs. This assumption, however, can be problematic in many situations, because a firm's technical and environmental efficiencies are expected to be correlated.

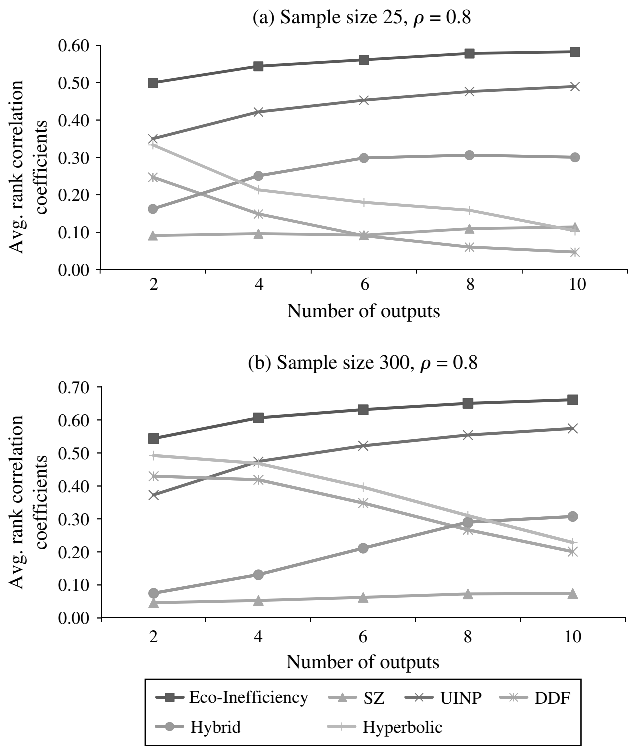
We develop a simulation framework for multiple desirable and undesirable outputs based on the concept from Fernández et al. (2002). However, we model the technical and environmental efficiencies as two correlated random variables. Specifically, we generalize the single-output function (11) to a multiple-output production function $F(X)$ of N desirable outputs $(y_1, \ldots, y_N)$ and P undesirable outputs $(u_1, \ldots, u_P)$ as

$$\begin{pmatrix} \log y_1 \\ \log y_2 \\ \vdots \\ \log y_N \\ \log u_1 \\ \log u_2 \\ \vdots \\ \log u_P \end{pmatrix} = F(X) + \boldsymbol{\nu} - \boldsymbol{\mu} = \log AX + \begin{pmatrix} \nu_1^y \\ \nu_2^y \\ \vdots \\ \nu_N^y \\ \nu_1^u \\ \nu_2^u \\ \vdots \\ \nu_P^u \end{pmatrix} - \begin{pmatrix} \mu_1^y \\ \mu_2^y \\ \vdots \\ \mu_N^y \\ -\mu_1^u \\ -\mu_2^u \\ \vdots \\ -\mu_P^u \end{pmatrix}, \tag{12}$$

where A denotes the coefficient matrix and each row of A has the log-linear structure of f in (11). As in the univariate production function, all random noise terms ν for different outputs in (12) follow an i.i.d. standard normal distribution. For the inefficiency effect, we distinguish between the *productive inefficiency* μ^y and the *environmental inefficiency* μ^u. The negative sign of the environmental inefficiency terms indicates that environmental inefficiency will cause firms to produce more undesirable outputs. Figure 5 in the online appendix illustrates the relationship between inputs, outputs, and the two inefficiency terms.

Specifically, μ^y and μ^u are the inefficiency effects associated with the production of desirable and undesirable outputs, respectively. The variable μ^y, the *productive inefficiency* term, is nonnegative and can reduce the desirable output quantities in $F(X)$. On the other hand, μ^u, the *environmental inefficiency* term, has the effect of increasing undesirable outputs quantities from the efficient level in $F(X)$. We assume that μ^y and μ^u are positively correlated (but not perfectly correlated). This is consistent with empirical findings from studies that show a significantly positive relationship between corporate environmental and

Figure 4. Kendall's tau under different numbers of outputs.



financial performance (e.g., Klassen and McLaughlin 1996, King and Lenox 2002).

Based on the assumption made in the conventional production function such as (11), we similarly assume the two inefficiency terms in the multivariate production function follow a bivariate half normal distribution: $(\mu^y, \mu^u) \sim |N_2(0, \Sigma)|$, where Σ is a semipositive definite variance-covariance matrix. The joint distribution function of (μ^y, μ^u) is (Kotz et al. 2002, pp. 326–327):

$$p(\mu^y, \mu^u) = \frac{2}{\pi\sigma_y\sigma_u\sqrt{1-\rho^2}}\exp\left(\frac{-(\mu^y/\sigma_y)^2-(\mu^u/\sigma_u)^2}{2(1-\rho^2)}\right) \cdot \cosh\left(\frac{\rho\mu^y\mu^u}{(1-\rho^2)\mu^y\mu^u}\right), \tag{13}$$

where σ_y and σ_u denote the standard deviation for μ^y and μ^u, respectively. Note that the marginal distribution μ^y, μ^u is half-normal, which matches the distributional assumption made in the univariate production function (10). The variance-covariance matrix Σ of a bivariate normal distribution can be written as a function of the standard deviations of (μ^y, μ^u) and the correlation coefficient ρ between (μ^y, μ^u) as (Gut 2009, p. 126):

$$\Sigma = \begin{bmatrix} \sigma_y^2 & \rho\sigma_y\sigma_u \\ \rho\sigma_y\sigma_u & \sigma_u^2 \end{bmatrix}. \tag{14}$$

This covariance structure allows us to vary the correlation between the productive and environmental efficiency terms by assigning different values to ρ in a simulation experiment. Based on Equation (12), the production function used in our experiment is given by

$$\begin{pmatrix} \log y_1 \\ \log y_2 \\ \vdots \\ \log y_N \\ \log u_1 \\ \log u_2 \\ \vdots \\ \log u_P \end{pmatrix} = \begin{pmatrix} \log\alpha_0 + \alpha_1\log x_1 + \alpha_2\log x_2 \\ \vdots \\ \log\alpha_0 + \alpha_1\log x_1 + \alpha_2\log x_2 \end{pmatrix} + \begin{pmatrix} \nu_1^y \\ \nu_2^y \\ \vdots \\ \nu_N^y \\ \nu_1^u \\ \nu_2^u \\ \vdots \\ \nu_P^u \end{pmatrix} - \begin{pmatrix} \mu_1^y \\ \mu_2^y \\ \vdots \\ \mu_N^y \\ -\mu_1^u \\ -\mu_2^u \\ \vdots \\ -\mu_P^u \end{pmatrix}. \tag{15}$$

To simplify the experimental setup, we let all outputs share the same coefficient values in the production function (15), but the output values (i.e., y_n and u_p) are contingent on the noise terms associated with each output (ν), as well as the productive or environmental inefficiency effects (i.e., μ^y and μ^u). Once we specify the input and the two stochastic terms in (15), we can calculate the output vector on the left-hand side of (15).

4.2. Evaluation Criteria

With the simulated input and output data, we can use the four frontier models discussed previously and our eco-inefficiency model and compute inefficiency scores. Comparing the inefficiency scores with the inefficiency variables in the simulation can indicate the performance of these frontier models. In this section we introduce two criteria, namely, *correlation* and *error rate*, which we use to evaluate the performance of the six frontier models.

4.2.1. Correlation Criterion. The validity of the frontier models hinges on how well the inefficiency scores correspond to the true inefficiency status of firms. To measure the validity of frontier models, we calculate the rank correlation between the inefficiency scores and the simulated inefficiency terms, which we operationalize as the inefficiency effect that frontier models are supposed to detect. We calculate rank correlation because inefficiency scores obtained from different frontier models may have their specific inefficiency indexes (see Table 1), and therefore rank correlation provides a more consistent assessment.

We expect that the rankings we derive from the inefficiency scores correlate highly with the "real" rankings, which we generate through simulation. Regarding the choice of correlation measures, we use the Kendall's tau (τ) rank correlation coefficient. Kendall's tau (τ) measures the degree of agreement between the generated and measured efficiency rankings. The tau (τ) coefficient ranges from -1 and 1, where "1" means a perfect match between two ranking distributions, and "-1" conversely suggests that one ranking distribution is the opposite of the other. See Kendall and Gibbons (1990) for an in-depth exposition of the Kendall's tau (τ) statistic.

In the production function (15), eco-inefficiency consists of productive inefficiency μ^y and environmental inefficiency μ^u. We use the average of these two inefficiency terms as the proxy of simulated eco-inefficiency ($\mu_{\text{avg}} = (\mu^y + \mu^u)/2$). Then we calculate the rank correlation coefficient between μ_{avg} and the inefficiency score θ obtained from a frontier model:

$$\tau = \text{corr}_{\text{kendall}}(\mu_{\text{avg}}, \theta). \tag{16}$$

4.2.2. Error Rate Criterion. In the paper mill example (online Appendix B), we observed that some mills' inefficiency scores decreased after we doubled their undesirable outputs. This is a clear indication of the issues raised by current frontier models, because the inefficiency score should be nondecreasing as the firm produces more pollution.

To measure the degree of inconsistency of frontier models, we record the number of times that the inefficiency score decreases (therefore the firm appears to be *less* eco-inefficient) after we experimentally double all undesirable outputs of the evaluated firm. When a firm's inefficiency score decreases in this situation, we call it *an error*. More specifically, we define the error rate for a frontier model as

$$\delta = \sum_{k=1}^{K} \omega_k \Big/ K \quad \text{where } \omega_k = 1 \text{ if } \theta_k^* - \tilde{\theta}_k^* > 0 \text{ and } \omega_k = 0 \text{ otherwise.} \tag{17}$$

In (17), θ_k^* stands for the inefficiency score of firm k obtained using the original data, whereas $\tilde{\theta}_k^*$ is the inefficiency score that we obtain from the same frontier model, but computed with the firm k's undesirable outputs doubled. Therefore, ω_k is equal to one when the firm k's score is an error, as defined earlier, and δ indicates the likelihood of an error in the sample.

4.3. Parameters

In our experiment, we control for four factors that could influence environmental efficiency estimates. These include sample size, number of inputs and outputs, correlation between productive and environmental inefficiencies, and returns-to-scale properties of the production function. Table 2 lists the simulation parameters in the experiment.

Table 2. Experiment parameters.

Parameters	Value
Sample size	$[25, 50, 100, 200, 300]$
Number of inputs	2
Number of desirable and undesirable outputs	2, 4, 6, 8, and 10
Correlation between the productive and environmental inefficiency terms (ρ)	$[0.2, 0.4, 0.8]$
Returns-to-scale parameters of the two-input Cobb-Douglas function	Increasing RTS: $\alpha_0 = 1$, $\alpha_1 = 0.65$, $\alpha_2 = 0.55$. Constant RTS: $\alpha_0 = 1$, $\alpha_1 = 0.5$, $\alpha_2 = 0.5$. Decreasing RTS: $\alpha_0 = 1$, $\alpha_1 = 0.35$, $\alpha_2 = 0.45$.
Probabilistic parameters	
Distribution of the input variable (x_1, x_2)	Uniform$[1, 4]$
Error term distribution (ν)	$N(0, 0.36)$
Inefficiency term distribution (μ)	$\lvert N(0, 5.06)\rvert$
Covariance matrix of the two inefficiency terms (Σ)	$\begin{bmatrix} 5.06 & \rho * 5.06 \\ \rho * 5.06 & 5.06 \end{bmatrix}$

Our choice of sample sizes is based on recent DEA simulation studies (Banker and Natarajan 2008, Kuosmanen and Johnson 2010). DEA-simulations have been mostly applied to small samples (Zhang and Bartels 1998, Adler and Yazhemsky 2010). This is because DEA is a nonparametric approach and is generally more robust to small samples than parametric approaches (e.g., the stochastic frontier model; Seiford and Thrall 1990). However, because DEA applications may involve larger samples, we include five different sample sizes in the experiment (25, 50, 100, 200, and 300), which correspond to small and large sample sizes in the applications of frontier models (Banker et al. 1993, Zhang and Bartels 1998). This allows us to test the performance of these different models with different sample sizes. We also vary the number of outputs to test whether the output dimensionality impacts the performance of these frontier models. In addition, we consider three sets of parameter values corresponding to increasing, constant, and decreasing RTS technologies.

Following Banker and Natarajan (2008), we generate the input variables x_1 and x_2 from a continuous uniform distribution between 1 and 4. We select three different values for the correlation parameter ρ between the productive and environmental inefficiency terms: a low correlation ($\rho = 0.2$), moderate correlation ($\rho = 0.4$), and high correlation ($\rho = 0.8$). In the simulation, we also test the performance of the model with more output variables. We do this by multiplying the number of output variables by two (i.e., two desirable and two undesirable outputs).

We follow prior simulation studies of frontier models and assume that the noise term has a standard normal distribution and the inefficiency term a half-normal distribution

(Pastor et al. 2002, Coelli et al. 2005, Greene 2005, Kuosmanen and Johnson 2010). We also assume that the productive and environmental inefficiency terms have a bivariate half-normal distribution, and that these two terms have equal variances for their marginal distributions. We designate the variance parameters for the inefficiency distribution as $\sigma_\nu^2 = 0.36$ and $\sigma_\mu^2 = 5.06$.[2] The above variance parameter values are chosen for two main reasons. First, to represent a realistic situation our experiment includes a moderate measurement error. The ratio between the variances of simulated inefficiency and the noise distributions is equal to 5.13, which, according to Banker and Natarajan (2008), corresponds to a situation with moderate measurement errors. Second, we also want the inefficiency score distributions that we obtain from the simulated input–output data to be realistic. By using the chosen variance values for the noise and inefficiency terms, the average eco-inefficiency score in our experiment is equal to 0.388, which is close to the average eco-inefficiency that we obtained in our prior evaluation of U.S. electric utility firms (Chen et al. 2010).[3]

4.4. Results

We replicate the simulation experiment 1,000 times under the parameter values shown in Table 2. In this simulation, we evaluate the frontier models with the average rank correlation coefficients τ and the consistency measure δ. We compare the average performance statistics of the five models under different sample sizes, inefficiency correlation coefficients ρ, and returns-to-scale assumptions of the multivariate Cobb-Douglas function. Full simulation results (Tables E.1 to E.3) can be found in the online Appendix E of the ecompanion for this paper (available at http://dx.doi.org/10.1287/opre.1120.1094).

4.4.1. Result 1: Rank Correlations. We now examine the correlation criterion that corresponds to the rank correlation (τ) between inefficiency scores and simulated inefficiencies. We first examine the influence of RTS assumptions and the inefficiency correlation parameters ρ on the rank correlation coefficients (τ). Note that these two parameters are usually exogenous factors in most applications of frontier models, so it is important to examine their effects on different frontier models. Regarding the effect of RTS assumptions on the correlation coefficients τ, we pool the τ coefficients of individual models obtained under different parameter settings (i.e., sample size, inefficiency correlation parameter ρ correlation, and number of outputs) and calculate summary statistics under three designated RTS parameters. Therefore, the τ coefficients associated with one particular set of parameters is considered a sample; i.e., for each model, we obtain $3 * 5 * 5 = 75\tau$ coefficient values, where the "3," "5," and "5" correspond to the experimental levels of ρ values, number of outputs, and sample sizes; see Table 2 for further details. The statistics are summarized in Table E.1 in the online appendix. Table E.1 shows that the τ coefficients are robust to different RTS settings and inefficiency correlation parameters ρ. For each model, we applied the nonparametric Kruskal-Wallis test to see if there are significant differences in the τ coefficients under different RTS assumptions (i.e., compare three populations with 75 paired samples). The Kruskal-Wallis test is employed because the distributions of τ coefficients are highly skewed to the right. The test results suggest that the median τ coefficients of all models are not significantly different RTS assumptions at ($p < 10\%$). These results mean that the rankings derived from these frontier models are robust to technological assumptions on RTS, although the inefficiency scores of a firm may vary under different RTS assumptions.

Table E.1 also shows the rank correlations under three different inefficiency correlation parameters ρ (0.2, 0.4, and 0.8). The Kruskal-Wallis test result shows that rank correlations under low, medium, and high values are significantly different at the 1% significant level for all models. In particular, the rank correlation τ increases systematically across all models as the ρ changes from low to high values. A possible explanation is that, when the correlation between two inefficiency terms is high, the simulated observations tend to deviate from the frontier more evenly among different outputs and move towards the interior of the feasible output space. This may increase the likelihood that the inefficiency scores become more accurate estimates of the average inefficiencies μ_{avg} defined in §4.1.1. We tested the interactive effects between RTS and the inefficiency correlation factors; the regression results show that only ρ has a significant effect on inefficiency scores and no significant interaction effects between ρ effects and RTS. Results are provided in the online Appendix D of the ecompanion for this paper (available at http://dx.doi.org/10.1287/opre.1120.1094). As our results are consistent across different RTS conditions, we confine our subsequent discussions on τ coefficients to simulations results based on the constant RTS assumption.

Figure 4 displays the τ coefficient values under small ($n = 25$) and large samples ($n = 300$). The average increase in τ coefficients when the sample size increases from 25 to 300 is 12% for the eco-inefficiency model, 13% for the UINP model, 150% for the DDF model, and 80% for the hyperbolic model.

The simulation results show that the τ coefficients increase as the sample size increases except for the SZ and the hybrid models (see Table E.2 in the online Appendix E). Several studies have indicated the advantage of using larger sample sizes in the frontier analysis (Banker 1993, Grosskopf 1996). With a larger sample, the frontier model has a higher likelihood to get a finer estimation of the frontier. Therefore, a larger sample in general helps reduce errors when calculating the inefficiency score (e.g., Grosskopf 1996, Zhang and Bartels 1998). We can also expect that a frontier model can gain a higher τ coefficient under a larger sample. The sharp increase in the

τ coefficients for the DDF and the hyperbolic model suggests that the performance of these two models is more sensitive to sample sizes than the other models (see the above statistics). One possible reason is that a larger sample not only improves estimation of the efficiency frontier for the DDF and the hyperbolic model, but also mitigates the problem associated with the weak disposability assumption. For example, when the sample size is small, it is comparatively more likely to obtain a misspecified frontier in these two models, and the degree of misspecification can also be higher with a small sample (i.e., the "coverage" and "slope" of the misspecified frontier; see Figures 2 and 5)—with a large sample, there is higher likelihood that one can obtain a sample point(s) above the misspecified frontier (e.g., Figure 2), which can mitigate the influence of a misspecified frontier. The error rates that we obtained for these two models corroborate our speculations (see our discussion in the next section).

Unlike the other models, the SZ and the hybrid models show the opposite results in the presence of a large sample: the τ value drops by 38% for of the SZ model and 15% for the hybrid model when the sample size increases from 25 to 300. We offer one possible explanation. We noted earlier that the translation vector in the SZ and hybrid models depends on the maximal undesirable output quantities, and the sensitivity of the SZ model decreases as the translation vector grows in magnitude (see the discussion in the paper mill application in the online Appendix B). As we increase the sample size, the chance of having more extreme observation also increases, which may ultimately bring down the sensitivity and hence the τ coefficient value.

Figure 4 also shows the τ coefficient values for models with 2, 4, 6, 8, and 10 outputs. The eco-inefficiency model has the highest average τ value among all six models under small and large samples. The DDF and hyperbolic models have lower τ values for a higher number of outputs. This is because when there are more undesirable outputs in these two models, it is more difficult to find a feasible vector of z_k as in (1.5) to form a linear span $\sum_{k=1}^{K} z_k u_{kp}^*$ that contains u_{op}^* for all p, where u_{op}^* is the benchmark value computed according to models (6) or (7). As a result, increasing the number of outputs also increases the chance and impact of a misspecified frontier for these two models.

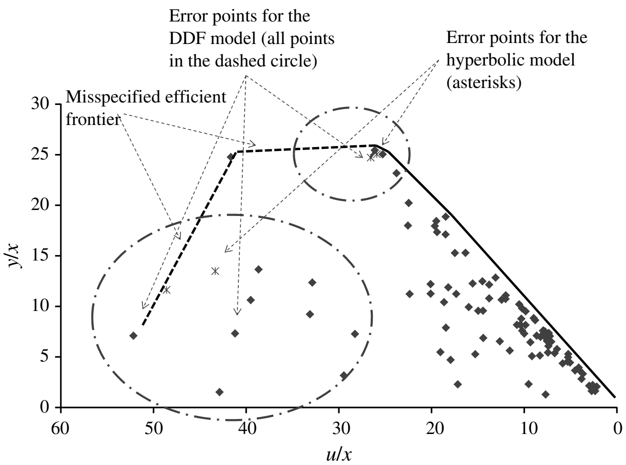
We also see from Figure 4 that the hybrid model, which utilizes the additive efficiency measure and the inverse translation technique from the SZ model, does not attain high correlation scores, compared with the scores of the eco-inefficiency model. We also observe that the performance of the hybrid model is consistently lower than the eco-inefficiency model across different settings. In addition to the influence of the translation vector mentioned earlier, another factor that helps explain the low performance can be seen from comparing the formulations of the hybrid and eco-inefficiency models. Specifically, we can rewrite the constraints for the undesirable outputs in the hybrid model as $\sum_{k=1}^{K} z_k(W_p - u_{kp}) \geqslant (W_p - u_{op}) + g_p^u$ for $p = 1, \ldots, P$, where W_p is the pth component of the translation vector W in the hybrid model. The above constraints can be recast as $\sum_{k=1}^{K} z_k u_{kp} \leqslant (u_{op} - g_p^u) + W_p(\sum_{k=1}^{K} z_k - 1)$ for $p = 1, \ldots, P$. The second set of inequality constraints resembles the constraints over undesirable outputs in our eco-inefficiency model, except that the right-hand side is now added by a variable term $(\sum_{k=1}^{K} z_k - 1)$ multiplied by W_p. Because the value of $\sum_{k=1}^{K} z_k$ depends on the relative economy of scales of the firm "o" in an industry (Banker et al. 1984), the inefficiency scores from the hybrid model depends on the scale economy factor, which may eventually lead to low correlation, as observed in our simulation experiment.

In summary, the simulation results indicate that our eco-inefficiency model outperforms other existing models in the correlation criterion. We also analyze how sample sizes and the number of outputs may influence the performance of the existing frontier models. We find that SZ and the hybrid models have lower τ values with a larger sample, and the DDF and hyperbolic models have lower τ values for a larger number of outputs.

4.4.2. Result 2: Error Rate Criterion. Now we turn to the δ values of the six models. The δ represents the likelihood that the inefficiency score of an observation decreases after we double the undesirable output quantities of this observation. The eco-inefficiency, SZ, hybrid, and UINP models exhibit zero δ values in the experiment, which means that we do not find any instances of errors for these four models without the weak disposability assumption. The hyperbolic and DDF models, however, show positive δ values. Table E.3 in the online Appendix E shows the δ scores of the DDF and hyperbolic models under different sample sizes, RTS, and inefficiency correlation ρ.

First, we find that the RTS factor does not significantly change the error rates ($p > 10\%$ in the Kruskal-Wallis test). Therefore, our results regarding the effect of RTS suggest that although RTS may change the inefficiency scores (and hence the imputed benchmark points), RTS does not have a significant impact on the efficiency rankings of firms. Second, we find that the error rates for both models tend to decrease with a larger sample. This finding is consistent with our results of the effect of sample sizes on correlation. Third, the average error rate becomes lower when we increase the number of outputs. Our conjecture is that with a higher output dimensionality, an observation is less probable to have its inefficiency score calculated based on the misspecified efficient frontier (see, e.g., line "*cd*" in Figure 1). Fourth, the simulation results confirm that problem of the weak disposability assumption that we observed with the analysis of the paper mill data. Although the average error rates from the experiment seem low in magnitude, we illustrate next that errors are much more likely to occur for firms that have relatively high amounts of desirable and undesirable outputs under the DDF and hyperbolic model.

Figure 5. Illustration of 100 simulated data points ($n = 100$, $\rho = 0.4$).



In Figure 5, we use the simulated data to illustrate the cause of errors in the DDF and hyperbolic models. For ease of illustration, we use the single-input Cobb-Douglas function from Banker and Natarajan (2008) to generate the data points. The figure contains 100 points of simulated desirable and undesirable outputs. Points that have led to errors under the DDF model are circled with the dashed line, and points that have yield errors under the hyperbolic model are marked with asterisks.

This figure illustrates the following points. First, points that have errors, as defined in §4.2.2, are those that identify benchmark targets on the misspecified efficient frontier when doubling undesirable outputs. Therefore, the more undesirable outputs a firm produces (compared with an average firm in the sample), the more likely a measurement error will occur for this firm under the DDF and the hyperbolic models. Second, the hyperbolic model is less prone to errors than the DDF model. The result echoes what we see in our simulation experiment (see Table E.3 in online Appendix E).

5. Discussion and Conclusion

As environmental awareness and pressure increases, there are pressing needs for managers and policymakers to use effective tools to assess corporate performance according to firms' input consumption, products, and undesirable outputs that could create negative externalities to the natural environment and society. However, undesirable outputs, such as greenhouse gas emissions or hazardous substances, usually do not have a fully functioning market that provides objective information about the relative importance of different factors. Consequently, aggregating multiple productive and ecological factors into a comprehensive index becomes a real challenge to both academics and practitioners.

In this paper, we develop a new nonparametric frontier model to evaluate a firm's eco-inefficiency. Our model allows us to construct the best-practice efficient frontier based on observed input–output quantities without the need to make explicit prioritization assumptions. Our model produces an eco-inefficiency score in the presence of multiple inputs and outputs. The eco-inefficiency score can help firms understand their competitive standing in their own industry and provide a concrete benchmark target for subsequent efficiency improvement activities.

Our paper makes major contributions to the frontier literature. We identify a fundamental issue associated with the weak disposability assumption on undesirable outputs in production economics. Specifically, we show that, under this assumption, existing frontier models may generate unreasonable estimations of eco-inefficiency scores and identify targets that are actually dominated in the production set. We compare our model with four alternative frontier models used in the literature. The results from the Monte Carlo experiment show that our approach provides a more robust measurement than these four frontier models. In the experiment, the eco-inefficiency model has attained higher rank correlations with the simulated inefficiency effect than the other models across all experimental conditions. We show that our eco-inefficiency model is guaranteed to identify eco-efficient points on the frontier, and therefore rectifies the inconsistency problem in efficiency measurement due to the weak disposability assumption on undesirable outputs in previous models. The simulation results confirm that the eco-inefficiency score is monotonic in undesirable outputs. The simulation model we employed also extends the traditional single-output production in the literature, which can only generate a single desirable output variable. We propose a new simulation framework amenable to the production process of multiple desirable and undesirable outputs. Our multioutput production function allows for greater flexibility and opens up a new path for the analysis of frontier models.

Our eco-inefficiency model has important implications for operations research and is not limited to the measurement of productive efficiency for operations involving environmental negative externalities. Indeed, many operations produce undesirable outcomes. These include accidents, delays, defective products, and waste. Our model can also be used for the measurement of efficiency frontiers in these situations.

Per definition, if outputs are undesirable, then the firm should seek to minimize them. Therefore we need an accurate frontier model that accommodates this. Carbon dioxide, along with other greenhouse gases, is still unregulated and is not priced in most markets. Without the price information, companies may resort to a quantity-based efficiency measure. In this case, the eco-inefficiency score is a quantity-based measure that indicates the evaluated firm's distance to the frontier.

One behavioral assumption behind our model is that firms are assumed to minimize their undesirable outputs.

The incentive to do so is clear if the emissions are regulated. However, there might be some other reasons for firms to reduce these emissions; for example, reputation effects and various benevolent side-effects such as improving production leanness. A good example is greenhouse gasses, which many of the largest firms are managing to curb, with the strong belief that mismanagement of the environmental practice can endanger corporate long-term sustainability (Delmas and Nairn-Birch 2011). More generally, we are seeing increasing evidence about the impact of firms' environmental records and stance toward sustainability on corporate performance. An important body of empirical literature shows that improved environmental performance leads to better corporate performance (see the following meta-analyses: Ambec and Lanoie 2008, King and Lenox 2002, Orlitzky et al. 2003). Increase in undesirable outputs can increase risk exposure (e.g., NGO or environmental activists' antagonistic campaigns, future legislations, failure to meet customers' environmental standards), which may eventually erode these companies' bottom lines. For example, greener suppliers are more likely to secure their market share because green suppliers reduce the buyer's environmental risk (Delmas and Montiel 2009). Some firms reduce their greenhouse gas emissions to reduce their risks, drive innovation, and save costs (Delmas and Nairn-Birch 2011).

We also point out some promising research directions. Because firms' environmental performance is receiving growing attention from market and governments, more firms are interested in the potential interactions between corporate eco-efficiency and different aspects of firm operations and management. The eco-inefficiency score provides an ideal proxy for eco-inefficiency to be used in empirical econometric models. See Banker and Natarajan (2008) for an updated procedure about how to regress inefficiency scores on independent variables of interests. One of the limitations of deterministic frontier models considered in this paper (as opposed to stochastic frontier models; see Coelli et al. 2005, Chapter 9, for an introduction) is that they do not consider the influence of statistical noise. As a result, the eco-inefficiency score may be sensitive to outliers in the sample or sampling errors. Therefore, a useful direction is to incorporate a stochastic term into the frontier formulation (e.g., Olesen and Petersen 1995, Post 2001, Post et al. 2002). Another promising direction is to carry out sensitivity analysis using bootstrapping (e.g., Simar and Wilson 1998) for the eco-inefficiency model. One might include stochastic price information in the eco-inefficiency model, in addition to capturing the statistical noise. In this paper we focus on the situation where output price information is unavailable, and we calculate the eco-inefficiency score based on the "quantities" of inputs and outputs. With complete price information, it is also possible to calculate the revenue efficiency of a firm when price information for output variables exists (see online Appendix F). One could also incorporate price into the objective function of the eco-inefficiency model, such that the inefficiency scores reflect the trade-off between undesirable and desirable outputs as signaled by prices. In practice, however, price information can be stochastic and can only be expressed in terms of probability distributions. For instance, greenhouse gas emissions have a price for companies operating under the E.U. carbon-trading scheme. A potential application would be to test how the carbon price fluctuation or pattern can affect firms' eco-efficiency.

Electronic Companion

An electronic companion to this paper is available as part of the online version at http://dx.doi.org/10.1287/opre.1120.1094.

Endnotes

1. From Google Scholar, Färe et al. (1989) receive 498 citations, Chung et al. (1997) receive 389 citations, Seiford and Zhu (2002) receive 209 citations, and Hailu and Veeman (2001) receive 105 citations (data retrieved February 28, 2012). These three papers also exhibit an increasing trend in the number of citations per year. Although studies citing these papers do not necessarily apply the original models, our intention in indicating the number of citations is to use that as an instrument to illustrate the persistent and growing influence of these four models.
2. The values of the variances of the two stochastic terms are taken from Banker and Natarajan (2008), where the distribution parameters are $\sigma_v = 0.04$ and $\sigma_u = 0.15$. We multiple these two values by 15 in order to obtain the desired average mean inefficiency, while maintaining a noise-to-signal ratio similar to that in Banker and Natarajan (2008).
3. Chen et al. (2010) evaluated the eco-inefficiency of 85 U.S. electric utility firms based on total sales (in MWH), three types of undesirable gases, and four inputs, and obtained an average eco-inefficiency score of 0.357 with a standard deviation 0.697. We choose the parameters in Table 4 such that we can obtain a higher sample average and standard division in the simulation with two inputs, six outputs, and sample size 100 (average 0.388; standard division 0.770). This is because prior studies have shown that including fewer input and output variables in a larger sample can lead to both higher sample average and variations for the inefficiency scores (Zhang and Bartels 1998).

Acknowledgments

The authors thank the associate editor and two anonymous referees for providing many constructive comments and suggestions that have vastly improved the focus and presentation of this article. They also thank Joe Zhu for sharing the paper mill data used in this paper.

References

Adler N, Yazhemsky E (2010) Improving discrimination in data envelopment analysis: PCA-DEA or variable reduction. *Eur. J. Oper. Res.* 202(1):273–284.

Ambec S, Lanoie P (2008) Does it pay to be green? A systematic overview. *Acad. Management Perspect.* 22(4):45–62.

Banker D, Natarajan R (2008) Evaluating contextual variables affecting productivity using data envelopment analysis. *Oper. Res.* 56(1):48–58.

Banker RD (1993) Maximum likelihood, consistency and data envelopment analysis: A statistical foundation. *Management Sci.* 39(10):1265–1273.

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

Banker RD, Gadh VM, Gott WL (1993) A Monte Carlo comparison of two production frontier estimation methods: Corrected ordinary least squares and data envelopment analysis. *Eur. J. Oper. Res.* 67(3):332–343.

Bardhan IR, Cooper WW, Kumbhakar SC (1998) A simulation study of joint uses of data envelopment analysis and statistical regressions for production function estimation and efficiency evaluation. *J. Productivity Anal.* 9(3):249–278.

Charnes A, Cooper W, Rhodes E (1978) Measuring the efficiency of decision making units. *Eur. J. Oper. Res.* 2:429–444.

Chen C-M (2012) Measuring environmental efficiency: Critical issues and solutions. Working paper, Nanyang Business School, Nanyang Technological University, Singapore.

Chen C-M, Delmas MA (2011) Measuring corporate social performance: An efficiency perspective. *Production Oper. Management* 20(6):789–804.

Chen C-M, Delmas MA, Montes-Sancho M (2010) Eco- vs. productive efficiency: A new approach to effective and comparative performance analysis. Working paper, Institute of the Environment and Sustainability, University of California, Los Angeles.

Chung YH, Färe R, Grosskopf S (1997) Productivity and undesirable outputs: A directional distance function approach. *J. Environ. Management* 51(3):229–240.

Coelli T, Rao P, O'Donnell CJ, Battese GE (2005) *An Introduction to Efficiency and Productivity Analysis* (Springer, New York).

Cooper WW, Park KS, Pastor JT (1999) RAM: A range adjusted measure of inefficiency for use with additive models, and relations to other models and measures in DEA. *J. Productivity Anal.* 11:5–42.

Cooper WW, Seiford LM, Tone K (2007) *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software* (Springer, New York).

Corbett CJ, Klassen RD (2006) Extending the horizons: Environmental excellence as key to improving operations. *Manufacturing Service Oper. Management* 8(1):5–22.

Corbett CJ, Kleindorfer PR (2001) Environmental management and operations management: Introduction to part 1 (manufacturing and ecologistics). *Production Oper. Management* 10(2):107–111.

Delmas MA, Blass VD (2010) Measuring corporate environmental performance: The trade-offs of sustainability ratings. *Bus. Strategy Environment* 19:245–260.

Delmas MA, Montiel I (2009) Greening the supply chain: When is customer pressure effective? *J. Econom. Management Strategy* 18(1):171–201.

Delmas MA, Nairn-Birch NS (2011) Is the tail wagging the dog? An empirical analysis of corporate carbon footprints and financial performance. Institute of the Environment and Sustainability, University of California, Los Angeles, Los Angeles. Accessed September 18, 2012, http://escholarship.org/uc/item/3k89n5b7.

Delmas MA, Toffel M (2008) Organizational responses to environmental demands: Opening the black box. *Strategic Management J.* 29(10):1027–1055.

Färe R, Grosskopf S (2004) *New Directions: Efficiency and Productivity* (Springer, New York).

Färe R, Grosskopf S, Lovell CAK, Pasurka C (1989) Multilateral productivity comparisons when some outputs are undesirable: A nonparametric approach. *Rev. Econom. Statist.* 71(1):90–98.

Färe R, Grosskopf S, Noh DW, Weber W (2005) Characteristics of a polluting technology: Theory and practice. *J. Econometrics* 126(2):469–492.

Fernández C, Koop G, Steel M (2002) Multiple-output production with undesirable outputs: An application to nitrogen surplus in agriculture. *J. Amer. Statist. Assoc.* 97:432–442.

Gibbins J, Chalmers H (2008) Carbon capture and storage. *Energy Policy* 36(12):4317–4322.

Golany B, Tamir E (1995) Evaluating efficiency-effectiveness-equality trade-offs: A data envelopment analysis approach. *Management Sci.* 41(7):1172–1184.

Greene W (2005) Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *J. Econometrics* 126(2):269–303.

Grosskopf S (1996) Statistical inference and nonparametric efficiency: A selective survey. *J. Productivity Anal.* 7:161–176.

Gut A (2009) *An Intermediate Course in Probability* (Springer, New York).

Hailu A, Veeman TS (2001) Non-parametric productivity analysis with undesirable outputs: An application to the Canadian pulp and paper industry. *Amer. J. Agricultural Econom.* 83(3):605–616.

Hua Z, Bian Y (2007) DEA with undesirable factors. Zhu J, Cook W, eds. *Modeling Data Irregularities and Structural Complexities in Data Envelopment Analysis* (Springer, New York), 103–122.

Kendall M, Gibbons JD (1990) *Rank Correlation Methods* (Oxford University Press, New York).

King AA, Lenox MJ (2002) Exploring the locus of profitable pollution reduction. *Management Sci.* 48(2):289–299.

Klassen R, Vachon S (2003) Collaboration and evaluation in the supply chain: The impact on plant-level environmental investment. *Production Oper. Management* 12(3):336–352.

Klassen RD, McLaughlin CP (1996) The impact of environmental management on firm performance. *Management Sci.* 42(8):1199–1214.

Koop G (1998) Carbon dioxide emissions and economic growth: A structural approach. *J. Appl. Statist.* 25(4):489–515.

Kotz S, Balakrishnan N, Johnson NL (2002) *Continuous Multivariate Distributions: Models and Applications* (John Wiley & Sons, New York).

Kuosmanen T, Johnson A (2010) Data envelopment analysis as nonparametric least-squares regression. *Oper. Res.* 58(1):149–160.

Olesen OB, Petersen NC (1995) Chance constrained efficiency evaluation. *Management Sci.* 41(3):442–457.

Orlitzky M, Schmidt FL, Rynes SL (2003) Corporate social and financial performance: A meta-analysis. *Organ. Stud.* 24(3):403–441.

Pastor JT, Ruiz JEL, Sirvent I (2002) A statistical test for nested radial DEA models. *Oper. Res.* 50(4):728–735.

Porter ME, Reinhardt F (2007) Grist: A strategic approach to climate. Forethought. *Harvard Bus. Rev.* 85(10):22–26.

Post T (2001) Performance evaluation in stochastic environments using mean-variance data envelopment analysis. *Oper. Res.* 49(2):281–292.

Post T, Cherchye L, Kuosmanen T (2002) Nonparametric efficiency estimation in stochastic environments. *Oper. Res.* 50(4):645–655.

Sahoo B, Luptacik M, Mahlberg B (2011) Alternative measures of environmental technology structure in DEA: An application. *Eur. J. Oper. Res.* 215(1):750–762.

Seiford LM, Thrall RM (1990) Recent developments in DEA. The mathematical programming approach to frontier analysis. *J. Econometrics* 46(1):7–38.

Seiford LM, Zhu J (2002) Modeling undesirable factors in efficiency evaluation. *Eur. J. Oper. Res.* 142(1):16–20.

Shephard RW (1970) *Theory of Cost and Production Functions* (Princeton University Press, Princeton, NJ).

Simar L, Wilson P (1998) Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Management Sci.* 44(1):49–61.

Tone K (2001) A slacks-based measure of efficiency in data envelopment analysis. *Eur. J. Oper. Res.* 130(3):498–509.

Zhang Y, Bartels R (1998) The effect of sample size on the mean efficiency in DEA with an application to electricity distribution in Australia, Sweden and New Zealand. *J. Productivity Anal.* 9(3):187–204.

Chien-Ming Chen is an assistant professor of operations management at the Nanyang Business School of Nanyang Technological University in Singapore. His research interests include environmental and sustainability issues in operations management, empirical research in operations and management, and theories and applications of production economics models.

Magali A. Delmas is a professor of management at the UCLA Institute of the Environment and the Anderson School of Management. She has published widely on business strategy and the natural environment. She studies information policies and strategies as they relate to sustainability. She is particularly interested in understanding the effectiveness of eco-labeling and certification strategies. She is engaged in refining current methodologies to measure and communicate firm and product environmental and social performance.