

On modeling pollution-generating technologies.

Sushama Murty* and R. Robert Russell**

July 2010

*Department of Economics, University of Warwick: s.murty@warwick.ac.uk

**Department of Economics, University of California, Riverside: robert.russell@ucr.edu

Abstract

We distinguish between intended production (of “goods”) and residual generation (of “bads”) and introduce the concept of by-production. We show that by-production provides the fundamental explanation for the commonly observed (or assumed) positive correlation between intended production and residual generation. Most of the existing literature attributes the observed positive correlation to abatement options available to firms. We show that abatement options of firms add to the phenomenon of by-production in strengthening the this positive correlation. The existing literature usually does not explicitly model abatement options of firms, but considers a reduced form of the technology that satisfies standard disposability assumptions with respect to all inputs and intended outputs. We show that more than one implicit production relation is needed to capture all the technological trade-offs that are implied by by-production. From our model, we are able to derive a reduced form of the technology that is in the spirit of the one that is usually studied in the literature. We find, however, that our reduced form technology violates standard disposability assumptions with respect to inputs and intended outputs that cause pollution. We derive implications from the phenomenon of by-production for econometric and Data Envelopment Analysis (DEA) specifications of pollution-generating technologies. We derive a DEA specification of technologies that satisfy by-production and employ it to measure technical efficiency of firms.

Journal of Economic Literature Classification Number: D20, D24, D62, Q50

Keywords: pollution-generating technologies, free disposability, weak disposability, data envelopment analysis, technical efficiency measurement.

On modeling pollution-generating technologies.

by

Sushama Murty and R. Robert Russell

1. Introduction.

Our reading of the environmental economics literature reveals three broad features of pollution that economists aim to capture. First, the generation of pollution/residuals seems to proceed hand-in-hand with the processes of consumption and production.¹ Second, the residuals so generated require the use of the assimilative capacity of the environment for their disposal. Third, the generation of the residuals and the consequent use of environmental resources for their disposal generate external effects on both consumers and producers and hence the need for policies to regulate the generation of pollution.

In this paper, we confine ourselves to addressing the first feature alone.² In particular, we focus on pollution generated by firms. We distinguish between outputs that firms intend to produce and outputs that unintentionally (incidentally) get generated by firms when they engage in the production of intended outputs. Pollution is such an unintended output. We are mainly concerned with studying the specification of technology sets that best captures the link between production of outputs intended by firms and the generation of pollution.

It is reasonable to say that, in the case of pollution generated by firms, there are some specific aspects about the process of transformation of inputs into intended outputs (*e.g.*, the use of certain inputs such as coal or the production of certain outputs such as varieties of cheese that release a strong odor) that trigger additional reactions in nature and (abstracting from abatement activities) *inevitably* result in the generation of pollution as a by-product. In this paper, we refer to these natural reactions, which occur alongside intended production by firms, as *by-production*³ of pollution.

In the case of technologies exhibiting by-production, we observe an inevitability of a certain *minimal* amount of the incidental output (the by-product), given the quantities of certain inputs and/or certain intended outputs. Inefficiencies in production could generate more than this minimal amount of the unintended output. At the same time, in such technologies, we also observe the usual menu of *maximal* possible vectors of intended outputs, given an input vector. Such a menu generally reflects the negative tradeoffs in the production

¹ See, especially, Ayres and Kneese [1969] and Førsund [2009].

² See Murty [2010] for a general equilibrium study of the second feature in the light of the first feature.

³ A word that is not in the dictionary but perhaps should be.

of intended outputs when inputs are held fixed, as production of each of these commodities is costly in terms of the inputs used. Inefficiencies in intended production may imply that less than this maximal amount may get produced. An increase in the amounts of the inputs used increases the menu of intended output vectors that are technologically feasible. At the same time, it increases the minimal amount of the unintended output that can be generated.⁴

The above underscores two crucial points to note about pollution-generating technologies:

- (i) technologies of pollution-generating firms do not satisfy free disposability of by-products such as pollution (pollution cannot be disposed of below the minimal level described above if inputs and intended outputs are held fixed) and
- (ii) in such technologies there is a mutual interdependence between changes in inputs, intended outputs, and pollution—an interdependence that we will argue is more correlation than causation.

In most of the existing literature, the standard building block employed in constructing pollution-generating technologies is the positive correlation between intended and unintended outputs that is usually observed in such technologies. This literature attributes this observed positive correlation to abatement activities by firms rather than directly to the phenomenon of by-production. Abatement activities of firms involve a diversion of their resources (inputs) to mitigate or clean up the pollution they produce.⁵ The production of these abatement activities is hence costly, given fixed amounts of resources: the more resources are diverted to abatement activities, the less they are available for producing intended outputs. Hence, an increase in the level of abatement activities leads concomitantly to both lower residual generation and lower production of intended output.

In this literature, however, abatement activities are not explicitly modeled as another set of outputs produced by firms.⁶ Rather, what is proposed is a “reduced form” of the technology in the space of inputs, by-products, and intended outputs. Special assumptions are made to ensure that such a technology exhibits a positive correlation between by-products and intended outputs, which is implicitly explained by abatement options open to firms. At

⁴ *E.g.*, a greater amount of usage of coal increases the quantity of both smoke and electricity generated.

⁵ In this paper, we model abatement activities as outputs of the firm. Examples are end-of-pipe treatment plants (that treat and clean water to remove the pollutant) and production of outputs like scrubbers (which reduce sulphur emissions). We abstract from long-run abatement options of development, purchase, and installation of new technologies that generate less pollution. See *e.g.*, Barbera and McConnell [1998], where abatement activities include both a purchase of abatement capital and a diversion of some amounts of the usual inputs of a firm towards running of the abatement capital.

⁶ For an exception, see Barbera and McConnell [1998].

the same time, it is also assumed that the technology satisfies the standard disposability assumptions with respect to *all* inputs and intended outputs. The approaches taken in the literature to model the positive correlation include: (a) treating pollution as a standard input,⁷ or (b) treating pollution as an output but with the technology satisfying the assumptions of weak disposability and null-jointness with respect to intended and unintended outputs.⁸ In empirical works, both parametric and non-parametric specifications of such technologies are often employed for measuring technical efficiency, marginal abatement cost, productivity, and growth when economic units also produce incidental outputs like pollution. Both Data Envelopment Analysis (DEA)⁹ and econometric approaches are employed in this literature.¹⁰

We propose a model of pollution-generating technologies that captures the salient features (i) and (ii) of the phenomenon of by-production identified above. Even without any reference to explicit abatement efforts by firms, the model generates a positive correlation between pollution generation and intended outputs. This is because the model recognizes and subsumes a residual generation mechanism that is set in motion when firms undertake production of intended goods. The residual generation mechanism is a relationship between pollution and commodities that cause pollution. If we assume that it is some inputs (*e.g.*, coal) that cause pollution, then an increase in these inputs causes an increase in pollution. At the same time, an increase in these inputs results also (under standard assumptions) in an increase in intended outputs (say electricity). Thus, the positive correlation between by-products and intended outputs exists, even in the absence of abatement activities.

We show that abatement options available to firms can also be explicitly factored into our model. When they are available, they form a part of both the production of intended

⁷ See, *e.g.*, Baumol and Oates [1988] and Cropper and Oates [1992]. (Unlike Cropper and Oates [1992], we abstract from the distinction between pollution and emissions.)

⁸ A technology satisfies weak disposability of intended and unintended outputs if the latter can be disposed of only in strict tandem with the disposition of the former, and it satisfies null jointness if zero pollution implies all intended output quantities are zero as well. See Section 4 for formal definitions of these concepts.

⁹ DEA is a mathematical programming approach to the construction of data-based technologies and the concomitant calculation of technological efficiency of individual firms (or other organizations). See Färe, Grosskopf, and Lovell [1994] for a basic description of DEA and Fried, Lovell, and Schmidt [2008] for surveys of recent developments.

¹⁰ For measurement issues based on parametric specifications of a technology that treats by-products as outputs and employs weak disposability and null jointness see, *e.g.*, Pittman [1983], Färe, Grosskopf, Noh, and Yaisawarng [1993], Coggins and Swinton [1994], Hailu and Veeman [1999], Murty and Kumar [2002, 2003], and Murty, Kumar, and Paul [2006]. For non-parametric set-theoretic approaches under similar assumptions on the technology see, *e.g.*, Färe, Grosskopf, and Pasurka [1986], Färe, Grosskopf, Lovell, and Pasurka [1989], Färe, Grosskopf, Noh, and Weber [2005], and Boyd and McClelland [1999]. See Zhou and Poh [2008] for a comprehensive survey of over a hundred papers employing this approach to the modeling of pollution-generating technologies.

output (as their production is also costly in terms of resources/inputs of the firm) and the residual generation mechanism (as they mitigate residual generation). From the full technology, we derive a reduced form technology that is in the spirit of those studied in the usual literature. We find, however, that while our reduced form technology satisfies standard disposability properties with respect to inputs and outputs that do *not* affect pollution generation, the disposability properties of this technology with respect to abatement and commodities (*e.g.*, coal) that cause pollution are ambiguous. In general, this reduced form technology does *not* satisfy free disposability in these inputs and outputs, contrary to common assumptions in the literature. We also show how the proposed technology can be employed in empirical works concerned with measurement issues. In particular, we provide a DEA construct of our technology and carry out an empirical application for measuring technical efficiency employing the hyperbolic and directional-distance-function efficiency indexes that are found in the literature.¹¹

With a view to understanding the correct relations between outputs and inputs in the presence of outputs like pollution we show, in Section 2, that a single implicit relation between outputs and inputs is not rich enough to capture, simultaneously, all the trade-offs between commodities that are implied by the phenomenon of by-production. In Section 3, we propose a model of a pollution-generating technology in which these inconsistencies in trade-offs are resolved, regardless of whether or not abatement options are open to firms. We also derive a reduced form of our technology that is in the spirit of the one that is usually studied in the literature. In Section 4, we compare our (by-production) approach to modeling pollution-generating technologies with the standard approaches taken in the literature. In Section 5, we provide a DEA-based empirical application of our by-production approach for measuring technical efficiency. In this section we abstract from abatement efforts of firms. In Section 6, we extend our DEA formulation of technology to incorporate abatement efforts of firms and provide two numerical examples to compare the reduced-form DEA technologies derived from using the by-production approach to the usual approach in the literature based on weak disposability and null-jointness. We conclude in Section 7.

2. Single equation representation of pollution-generating technologies.

We show that a single implicit relation between outputs and inputs is not rich enough to capture, simultaneously, all the trade-offs between commodities that are implied by the phenomenon of by-production.

¹¹ See the literature cited in Footnote 10.

2.1. The case without abatement output.

The vectors of input quantities (indexed by $i = 1, \dots, n$), intended-output quantities (indexed by $j = 1, \dots, m$), and incidental-output quantities (indexed by $k = 1, \dots, m'$), are given, respectively, by $y \in \mathbf{R}_+^m$, $z \in \mathbf{R}_+^{m'}$, and $x \in \mathbf{R}_+^n$.

Suppose pollution is caused by the use of certain inputs like coal or because of the production of certain intended outputs like cheese. Suppose also that the firm does not participate in any abatement activity to reduce the pollution that it generates. A single equation formulation of such a pollution-generating technology, an extension of the standard functional representation of a multiple-output technology, is as follows:

$$T = \{ \langle x, y, z \rangle \in \mathbf{R}_+^{n+m+m'} \mid f(x, y, z) \leq 0 \},$$

where f is differentiable, with derivatives with respect to inputs and intended outputs given by

$$\begin{aligned} \text{(a)} \quad & f_i(x, y, z) \leq 0, \quad i = 1, \dots, n, \\ \text{(b)} \quad & f_j(x, y, z) \geq 0, \quad j = 1, \dots, m, \end{aligned} \tag{2.1}$$

where subscripts on f indicate partial differentiation with respect to the indicated variable.

The constraints (a) and (b) are standard differential restrictions to impose “free disposability” of, respectively, inputs and intended outputs:

$$\langle x, y, z \rangle \in T \wedge \bar{x} \geq x \implies \langle \bar{x}, y, z \rangle \in T \tag{2.2}$$

and

$$\langle x, y, z \rangle \in T \wedge \bar{y} \leq y \implies \langle x, \bar{y}, z \rangle \in T. \tag{2.3}$$

To capture the fact that pollution is an output of the production process for which disposal is not free, Murty [2010] introduces and formalizes an assumption that is the polar opposite of free output disposability with respect to the unintended outputs:

$$\langle x, y, z \rangle \in T \wedge \bar{z} \geq z \implies \langle x, y, \bar{z} \rangle \in T. \tag{2.4}$$

Following Murty [2010], we refer to this property as “costly disposability” of residuals.¹² The differential restriction required to impose costly disposability on T is

$$f_k(x, y, z) \leq 0, \quad k = 1, \dots, m'. \tag{2.5}$$

¹² Costly disposability implies the possibility of inefficiencies in the generation of pollution. If a given level of coal generates some level of smoke, then inefficiency in the use of coal may imply that this level of coal can also generate a greater amount of pollution. See also footnote 18 in Section 3.1.

Quantity vectors satisfying $f(x, y, z) = 0$ are points on the boundary of the technology. Those satisfying $f(x, y, z) < 0$ are inefficient: more intended output could be produced with given quantities of inputs and pollution; less pollution could be generated with given intended output and input quantities; and smaller input quantities could be used to produce the given output quantities, given the pollution level.¹³

Assume, in this section and without loss of generality, that $m' = 1$. Suppose $f_k(\hat{x}, \hat{y}, \hat{z}) < 0$ for some $\langle \hat{x}, \hat{y}, \hat{z} \rangle$ satisfying $f(\hat{x}, \hat{y}, \hat{z}) = 0$. Then, from the implicit function theorem, there exist neighborhoods $U \subseteq \mathbf{R}_+^{m+n}$ and $V \subseteq \mathbf{R}_+$ around $\langle \hat{x}, \hat{y} \rangle \in \mathbf{R}_+^{n+m}$ and $\hat{z} \in \mathbf{R}_+$ and a function $\zeta : U \rightarrow V$ such that

$$\hat{z} = \zeta(\hat{x}, \hat{y}) \quad (2.6)$$

and

$$f(x, y, \zeta(x, y)) = 0. \quad (2.7)$$

The trade-off between each intended output and unintended output k (with inputs and all other outputs held fixed) implied by the implicit function theorem is

$$\frac{\partial \zeta(x, y)}{\partial y_j} = -\frac{f_j(x, y, z)}{f_k(x, y, z)} \geq 0, \quad j = 1, \dots, m. \quad (2.8)$$

The trade-off between each input and unintended output k (with intended outputs and all other inputs held fixed) is

$$\frac{\partial \zeta(x, y)}{\partial x_i} = -\frac{f_i(x, y, z)}{f_k(x, y, z)} \leq 0, \quad i = 1, \dots, n. \quad (2.9)$$

Noting that all these trade-offs are evaluated at points in the technology set that are (weakly) technically efficient (that is, $f(x, y, z) = 0$), the foregoing formulation of a pollution-generating technology seems to be inconsistent with the phenomenon of by-production for the following reasons:

- (a) The existence of the function ζ satisfying (2.8) as a strict inequality implies that there exists a rich menu (a manifold) of (technically efficient) $\langle y, z \rangle$ combinations, with varying levels of z , that are possible with *given levels of all inputs*. If pollution is generated by input usage, then this menu is contrary to phenomenon of by-production, since the phenomenon implies that at fixed levels of inputs (*e.g.*, coal), there will be only *one* technically efficient (minimal) level of pollution.¹⁴

¹³ Of course, given the weak inequalities in the constraints (2.1) and (2.5), the set of efficient (frontier) points is a subset of the boundary of the technology.

¹⁴ If pollution is caused by some intended outputs (*e.g.*, strong odor from some varieties of cheese produced by a dairy) and (2.9) holds as a strict inequality, then it implies that there exists a rich menu of (technically efficient) $\langle x, z \rangle$ combinations, with varying levels of z , that are possible with *given levels of all intended outputs*. Such a menu is inconsistent with by-production.

(b) Furthermore, if pollution is generated by inputs such as coal, as is very often the case, the negative trade-offs between pollution generation and inputs (derived by holding the levels of intended outputs fixed), apparent in (2.9), are inconsistent with by-production, which implies that this trade-off should be positive.

How should one interpret the trade-offs observed under single equation modeling of pollution-generating technologies when one abstracts from abatement options? As discussed above, these trade-offs are not reflective of the phenomenon of by-production. Rather, the positive trade-offs observed in (2.8) between each intended output and pollution and the negative trade-offs observed in (2.9) between each input and pollution suggest that this approach treats pollution like any other input in production: first, increases in its level, holding all other inputs fixed, increases intended outputs and, second, pollution is a substitute for all other inputs in intended production—the same level of intended outputs can be produced by decreasing other inputs and increasing pollution. This also does not seem to be intuitively correct: it is not a correct description of the role pollution plays in intended production.¹⁵

2.2. The case with abatement output.

Consider the case where the technology of a pollution-generating firm is defined by a single restriction on all inputs and outputs, including the abatement output:

$$T = \{ \langle x, y, z, y^a \rangle \in \mathbf{R}^{n+m+m'+1} \mid f(x, y, z, y^a) \leq 0 \}. \quad (2.10)$$

We assume that

$$f_a(x, y, z, y^a) \geq 0. \quad (2.11)$$

This restriction captures the fact that the abatement output is also freely disposable:

$$\langle x, y, z, y^a \rangle \in T \wedge \bar{y}^a \leq y^a \implies \langle x, y, z, \bar{y}^a \rangle \in T, \quad (2.12)$$

so that producing it is costly in terms of input usage, implying a negative trade-off between it and the other intended outputs. In that case, the implicit function theorem can again be invoked to show that, the trade-off between the abatement output and pollution, evaluated in a neighborhood of a (technically efficient) point $\langle \hat{x}, \hat{y}, \hat{z}, \hat{y}_a \rangle \in \mathbf{R}_+^{n+m+m'+1}$ such that $f(\hat{x}, \hat{y}, \hat{z}, \hat{y}_a) = 0$ and $f_k(\hat{x}, \hat{y}, \hat{z}, \hat{y}_a) < 0$, is

$$\frac{\partial \zeta(x, y, y_a)}{\partial y^a} = - \frac{f_a(x, y, z, y_a)}{f_k(x, y, z, y_a)} \geq 0 \quad (2.13)$$

whenever $f(x, y, z, y_a) = 0$, contradicting the fact that abatement output is produced by firms to mitigate, and not to enhance, pollution.

¹⁵ See also footnote 25 in Section 4.

3. A by-production approach to modeling pollution.

Given the above analysis, a sound foundation must be identified for introducing multiple production relations to capture correctly the features of by-production. We feel that the resolution to the problem lies in early work of Frisch [1965] on production theory, in which he envisaged situations where the correct functional representation of a production technology may require more than one implicit functional relation between inputs and outputs. More recently, Førsund [2009] explores these ideas of Frisch.¹⁶ We build on the works of Frisch and Førsund and show that the phenomenon of by-production requires distinguishing explicitly the by-product generating mechanism from the production relation that describes the production of intended commodities. We show that when this is done the inconsistencies among trade-offs elucidated in Section 2 get resolved.

3.1. A by-production approach: the case without abatement.

In this sub-section, we abstract from explicit abatement efforts. The production of the intended output sets a residual-generation mechanism in motion, leading to the generation of the by-product. To fix our ideas on the salient aspects of by-production and to simplify notation, we continue to assume, without loss of generality, that $m' = 1$ and that the pollution is generated by usage of a single input (such an input could be coal), say input ι , and the production of one of the firm's intended outputs, say output j . Denote the input and output quantity vectors purged of the quantity of input ι and the quantity of the output j by x^1 and y^1 , respectively. Specify the technology as

$$T = T_1 \cap T_2, \quad (3.1)$$

where

$$T_1 = \{ \langle x^1, x_\iota, y^1, y_j, z \rangle \in \mathbf{R}^{n+m+1} \mid f(x^1, x_\iota, y^1, y_j) \leq 0 \}, \quad (3.2)$$

$$T_2 = \{ \langle x^1, x_\iota, y^1, y_j, z \rangle \in \mathbf{R}^{n+m+1} \mid z \geq g(x_\iota, y_j) \}, \quad (3.3)$$

and f and g are continuously differentiable functions. The set T_1 is a standard technology set, reflecting the ways in which the inputs can be transformed into intended outputs. The standard free disposability properties (2.3) and (2.4) can be imposed on this set by assuming that f satisfies

$$\begin{aligned} f_i(x, y) &\leq 0, \quad i = 1, \dots, n, \quad \text{and} \\ f_j(x, y) &\geq 0, \quad j = 1, \dots, m. \end{aligned} \quad (3.4)$$

¹⁶ He employs a welfare maximization problem to show that the optimal government policies are counter-intuitive and meaningless when a single production relation is used to represent a pollution-generating technology.

Note, (3.2) imposes no constraint on z , that is, it is implicitly assumed that the by-product does not affect the production of intended outputs.¹⁷

The set T_2 reflects nature's residual-generation mechanism. T_2 satisfies costly disposability of pollution as defined in (2.4), with the function g defining the minimal level of pollution that gets generated for given levels of x_i and y_j .¹⁸ The derivatives of g satisfy

$$\begin{aligned} g_i(x_i, y_j) &\geq 0, \\ g_j(x_i, y_j) &\geq 0 \end{aligned} \tag{3.5}$$

The conditions in (3.5) capture the fact that the efficient (minimal) level of pollution rises with the increase in the usage of input i or the production of the intended output j . This means, however, that T_2 violates standard free disposability of input i . In fact it satisfies the polar opposite condition in these goods.¹⁹

$$\langle x^1, x_i, y^1, y_j, z \rangle \in T_2 \wedge \bar{z} \geq z \wedge \bar{x}_i \leq x_i \wedge \bar{y}_j \leq y_j \implies \langle x^1, \bar{x}_i, y^1, \bar{y}_j, \bar{z} \rangle \in T_2. \tag{3.6}$$

It is easy to infer the disposability properties of T from the disposability properties of the intended production technology T_1 and the residual generation mechanism T_2

Theorem 1: *T satisfies free disposability with respect to all intended outputs and non-pollution causing inputs. However, it violates free-disposability with respect to the pollution causing input i . It satisfies costly disposability with respect to the quantity of pollution z .*

¹⁷ This could be generalized, of course, allowing pollution to have an effect on intended production as well; *e.g.*, smoke could adversely affect the productivity of labor engaged in producing intended outputs. In that case, suitable adjustments can be made to the analysis below to take account of this generalization. These adjustments may impinge on the disposability assumptions that the pollution-generating technology satisfies and on the trade-offs among various commodities.

¹⁸ Costly disposability, as defined in (2.4), could be considered to be too extreme. It implies that an infinite amount of pollution can be generated by given amounts of input i and output j . In general, there may also be an upper bound for the generation of the unintended output. Let the set \mathcal{T}_2 reflect the realistic bounds on the generation of the by-product. Then $\mathcal{T}_2 \subseteq T_2$ and both sets have a common lower boundary defined by the function g (in fact, T_2 could be a particular monotonic hull of \mathcal{T}_2). From the point of view of technical efficiency and the econometric and DEA approaches for constructing technologies of pollution-generating firms and using the constructs for measurement issues, it is only this lower bound that is important. Hence, we focus only on the set T_2 .

¹⁹ This assumption reflects the possibility of inefficiencies in the production of pollution: if given levels of coal and a pollution-generating intended output generate some amount of pollution, then inefficiencies in residual generation may imply that lower amounts of the coal input or the intended output can generate the same level of pollution if the firm operates more efficiently.

The technology violates standard disposability conditions with respect to the quantity of the pollution-causing input x_i because, while T_1 satisfies standard free-disposability conditions in x_i , T_2 satisfies the polar opposite conditions with respect to this input.

Quantity vectors $\langle x, y, z \rangle \in T$ that satisfy $f(x, y) = 0$ and $z = g(x_i, y_j)$ are the weakly efficient points of T . If a quantity vector in $\langle x, y, z \rangle \in T$ is such that $f(x, y) < 0$, then it is technologically impossible to increase the levels of the non-pollution-causing intended outputs or decrease the levels of the non-pollution-causing inputs without changing the production levels of the remaining goods. If a quantity vector in $\langle x, y, z \rangle \in T$ is such that $z > g(x_i, y_j)$, it is technologically possible to decrease the level of pollution without changing the production levels of all other goods.

To sign the trade-offs between pollution and a non-pollution-causing intended output j at a frontier point of T , we invoke the implicit function theorem. Let $\langle \hat{x}, \hat{y}, \hat{z} \rangle$ be a frontier point of T . Then

$$\begin{aligned} f(\hat{x}, \hat{y}) &= 0 \\ \hat{z} - g(\hat{x}_i, \hat{y}_j) &= 0. \end{aligned} \tag{3.7}$$

Denote y^{-j} to be the vector obtained by purging the j^{th} element from the vector y , where $j \neq j$. Suppose that $f_j(\hat{x}, \hat{y}) \neq 0$ and $g_i(\hat{x}_i, \hat{y}_j) \neq 0$. Then the matrix

$$\begin{bmatrix} f_j(\hat{x}, \hat{y}) & f_i(\hat{x}, \hat{y}) \\ 0 & -g_i(\hat{x}_i, \hat{y}_j) \end{bmatrix} \tag{3.8}$$

has full row rank. By the implicit function theorem, there exists a neighborhood U around $\langle \hat{x}^1, \hat{y}^{-j}, \hat{z} \rangle$ in \mathbf{R}_+^{n+m} , a neighborhood V around $\langle \hat{x}_i, \hat{y}_j \rangle$ in \mathbf{R}_+^2 , and continuously differentiable mappings $\psi^j : U \rightarrow \psi^j(U)$ and $h : U \rightarrow h(U)$ with images

$$\begin{aligned} y_j &= \psi^j(x^1, y^{-j}, z) \\ x_i &= h(x^1, y^{-j}, z) \end{aligned} \tag{3.9}$$

such that $\langle h(x^1, y^{-j}, z), \psi^j(x^1, y^{-j}, z) \rangle \in V$ and

$$\begin{aligned} f(x^1, h(x^1, y^{-j}, z), \psi^j(x^1, y^{-j}, z), y^{-j}) &= 0 \\ z - g(h(x^1, y^{-j}, z), y_j) &= 0. \end{aligned} \tag{3.10}$$

In that case, assuming that $g_i(x_i, y_j) > 0$, the trade-off between y_j and z is

$$\frac{\partial \psi^j(x^1, y^{-j}, z)}{\partial z} = - \frac{f_i(x, y) h_k(x^1, y^{-j}, z)}{f_j(x, y)} \geq 0. \tag{3.11}$$

How should one interpret this non-negative ‘‘trade-off’’ between y_j and z seen in (3.11)? Starting at a weakly inefficient point in a local neighborhood of $\langle \hat{x}, \hat{y}, \hat{z} \rangle \in T$, an increase in

z is attributable, because of the by-production phenomenon inherent in T_2 , to an increase in x_i if y_j is held fixed (as $h_k(x^1, y^{-j}, z) > 0$).²⁰ Under the conventional assumptions on intended production in (3.4), the trade-off between the pollution-generating input i and intended output j is

$$-\frac{f_i(x, y)}{f_j(x, y)}; \geq 0, \quad (3.12)$$

hence, the increase in x_i implies an increase in y^j . The “trade-off” in (3.11), thus, reflects a non-negative *correlation* between the residual and an intended output via x_i , because a change in x_i affects both y^j (non-negatively in intended production) and z (positively with respect to residual generation); this is not a trade-off in the usual economic sense.

To summarize, the non-negative “trade-off” between an intended and an unintended output in the reduced form model is explained by (a) the phenomenon of by-production, which relates the use of inputs such as i to the by-product, and (b) the non-negative marginal product of input i in producing intended outputs like j .

3.2. A by-production approach: incorporating abatement activities.

We again keep the analysis simple by sticking to a single abatement output (as well as a single unintended output). On the other hand, we make the model more general to allow the possibility of input substitutability in the generation of the by-product.²¹ We do so by partitioning the vector of all n inputs into n_1 non-residual-generating inputs and n_2 residual-generating inputs. Denote the respective input quantity vectors by x^1 and x^2 . Let y^a denote the level of the firm’s abatement activities, which are also costly in terms of the input resources of the firm. Without loss of generality, we assume that the intended outputs do not cause pollution. Similarly to the previous section, we specify the technology as $T = T_1 \cap T_2$, where

$$\begin{aligned} T_1 &= \{ \langle x^1, x^2, y, z, y^a \rangle \in \mathbf{R}^{n+m+2} \mid f(x^1, x^2, y, y^a) \leq 0 \} \\ T_2 &= \{ \langle x^1, x^2, y, z, y^a \rangle \in \mathbf{R}^{n+m+2} \mid z \geq g(x^2, y^a) \}. \end{aligned} \quad (3.13)$$

T reflects both the transformation of inputs into intended outputs and abatement output (as indicated by the definition of T_1) and the use of the abatement output by the firm to control the by-production of the residual that results from use of pollution-generating inputs in producing intended outputs (as indicated by the definition of T_2 in (3.13)). We confine

²⁰ Note that the function h is the inverse of g : $h(x^1, y^{-j}, z) = g^{-1}(z, y_j)$, so that, if $z = g(x_i, y_j)$ and $g_i(x_i, y_j) > 0$, then the derivative of h with respect to z is $h_k(x^1, y^{-j}, z) = 1/g_i(x_i, y_j) > 0$.

²¹ For example, substituting a cleaner variety of coal for a less pure variety or vice-versa.

ourselves again to a local analysis and posit the following signs of the partial derivatives at a weakly efficient point $\langle \hat{y}, \hat{y}_a, \hat{x}^1, \hat{x}^2, \hat{z} \rangle$ of T :

$$\begin{aligned}
f_j(\hat{x}^1, \hat{x}^2, \hat{y}, \hat{y}_a) &\geq 0, \quad j = 1, \dots, m, \\
f_a(\hat{x}^1, \hat{x}^2, \hat{y}, \hat{y}_a) &> 0, \\
f_i(\hat{x}^1, \hat{x}^2, \hat{y}, \hat{y}_a) &\leq 0, \quad i = 1, \dots, n, \\
g_a(\hat{x}^2, \hat{y}_a) &< 0, \\
g_i(\hat{x}^2, \hat{y}_a) &\geq 0 \quad \text{for all } i = n_1 + 1, \dots, n, \\
g_i(\hat{x}^2, \hat{y}_a) &> 0 \quad \text{for some } i = n_1 + 1, \dots, n.
\end{aligned} \tag{3.14}$$

It is easy to see that (3.13) and (3.14) imply that T_1 satisfies standard free disposability conditions for inputs, abatement output, and intended outputs. In addition, there is a negative (or at least non-positive) trade-off between standard outputs and the abatement output and a positive (or a non-negative) trade-off between each intended output and the inputs in intended production.

With respect to residual generation, (3.13) and (3.14) imply that T_2 satisfies costly disposability for the unintended output and a condition that is polar opposite of standard input and output free disposability for the abatement output and non-pollution-generating inputs:

$$\langle x^1, x^2, y, z, y^a \rangle \in T_2 \wedge \bar{z} \geq z \wedge \bar{x}^2 \leq x^2 \wedge \bar{y}^a \geq y^a \implies \langle x^1, \bar{x}^2, y, \bar{z}, \bar{y}^a \rangle \in T_2. \tag{3.15}$$

We call (3.15) ‘‘costly disposability of pollution, abatement output, and inputs that generate pollution.’’²² The trade-offs between z and each of the pollution-generating input quantities x_i^2 implied by (3.14) are non-negative and that between z and abatement output y^a is negative. Thus, the sign of g_a captures the mitigating effect abatement has on residual generation and the sign of g_i captures the increase in pollution attributable to the increase in inputs causing pollution.

It is easy to infer the disposability properties of T from the above characteristics of T_1 and T_2 :

²² This assumption reflects the inefficiencies in the production of pollution: if given levels of coal and abatement activities generate some amount of pollution, then inefficiencies in the use of coal or abatement activities imply that a lower amount of the coal input or a higher level of abatement activities could generate the same level of pollution if the firm operates more efficiently.

Theorem 2: *T satisfies free disposability with respect to all intended outputs and non-pollution-causing inputs. However, it violates free disposability with respect to each of the pollution-causing inputs and the abatement output. It satisfies costly disposability with respect to pollution.*

Let the inequalities in (3.14) hold. We now sign the trade-off between z and an intended output y_j at a weakly efficient point of T . As in the previous section, we do so by employing the implicit function theorem. Let $\langle \hat{x}^1, \hat{x}^2, \hat{y}, \hat{z}, \hat{y}^a \rangle$ be a weakly efficient point of T . Then

$$\begin{aligned} f(\hat{x}^1, \hat{x}^2, \hat{y}, \hat{y}^a) &= 0 \\ \hat{z} - g(\hat{x}^2, \hat{y}^a) &= 0. \end{aligned} \tag{3.16}$$

Let $f_j(\hat{x}^1, \hat{x}^2, \hat{y}, \hat{y}^a) \neq 0$ and $g_a(\hat{x}^2, \hat{y}^a) \neq 0$. Then the matrix

$$\begin{bmatrix} f_j(\hat{x}^1, \hat{x}^2, \hat{y}, \hat{y}^a) & f_a(\hat{x}^1, \hat{x}^2, \hat{y}, \hat{y}^a) \\ 0 & -g_a(\hat{x}^2, \hat{y}^a) \end{bmatrix} \tag{3.17}$$

is full-row ranked. The implicit function theorem implies that there exists a neighborhood U around $\langle \hat{x}, \hat{y}^{-j}, \hat{z} \rangle$ in \mathbf{R}_+^{n+m} , a neighborhood V around $\langle \hat{y}^j, \hat{y}^a \rangle$ in \mathbf{R}_+^2 , and continuously differentiable mappings $\psi^j : U \rightarrow \psi^j(U)$ and $h : U \rightarrow h(U)$ with images

$$\begin{aligned} y_j &= \psi^j(x, y^{-j}, z) \\ y^a &= h(x, y^{-j}, z) = g^{-1}(z, x^2) \end{aligned} \tag{3.18}$$

such that $\langle \psi^j(x, y^{-j}, z), h(x, y^{-j}, z) \rangle \in V$ and

$$\begin{aligned} f(x, \psi^j(y^{-j}, z), y^{-j}, h(x, y^{-j}, z)) &= 0 \\ z - g(x^2, h(x, y^{-j}, z)) &= 0. \end{aligned} \tag{3.19}$$

In that case, the trade-off between y_j and z is

$$\frac{\partial \psi^j(x, y^{-j}, z)}{\partial z} = - \frac{f_a(x, y, y^a) h_k(x, y^{-j}, z)}{f_j(x, y, y^a)} \geq 0. \tag{3.20}$$

As in the previous section, this non-negative trade-off between an intended output and pollution at a weakly efficient point of T reflects a correlation between these commodities; in this case, this correlation is effected by abatement effort of the firm to mitigate by-production of pollution.²³ Precisely, holding the levels of all inputs (including pollution-causing inputs) fixed, an increase in z must have come about because of reductions in abatement efforts y^a

²³ Note that, as in the previous section, a (generally different) positive correlation between the intended and unintended outputs effected by an input that causes pollution could also be derived.

by firms, and hence there is an increase in resources diverted towards production of other intended outputs y (assuming, of course, that firms are operating efficiently).

From our analysis above, we can derive the reduced-form functional representation of the technology T . By substituting out abatement efforts from the function f in (3.13), we can rewrite T equivalently as

$$T = \{ \langle x^1, x^2, y, z, y^a \rangle \in \mathbf{R}^{n+m+2} \mid \tilde{f}(x, y, z) \leq 0 \wedge y^a \geq h(x, y^{-j}, z) \}, \quad (3.21)$$

where

$$\tilde{f}(x, y, z) := f(x, y, h(x, y^{-j}, z)). \quad (3.22)$$

Using (3.21), we can define a reduced-form technology in the space of intended and unintended outputs and inputs as

$$\tilde{T} := \{ \langle x^1, x^2, y, z \rangle \in \mathbf{R}_+^{n+m+1} \mid \tilde{f}(x, y, z) \leq 0 \}. \quad (3.23)$$

It is easy to check that, in the neighborhood of a weakly efficient point $\langle x, y, z, y^a \rangle$ of T , the trade-off between an intended and an unintended output, $-\tilde{f}_j(x, y, z)/\tilde{f}_k(x, y, z)$, is given by (3.20) and hence is non-negative.

4. A comparison of conventional formulation of a pollution-generating technology and the by-production approach.

As discussed in the introductory section, the conventional literature models a reduced-form technology—quite in the spirit of \tilde{T} (in (3.23) above)—in the space of intended and unintended outputs and inputs that exhibits a positive correlation between intended and unintended outputs but satisfies *all* of the standard free disposability assumptions with respect to intended outputs and inputs. The technology is modeled only in reduced form because, although this literature attributes the positive correlation to abatement options available to firms, abatement activities are not explicitly modeled.

In this literature, pollution is treated either as a standard input (specifying a function decidedly in the spirit of \tilde{f} in Section 3.2, with derivatives satisfying sign restrictions (2.1) and (2.5)) or is treated as an output (with novel disposability conditions). In the latter case, the reduced-form technology is represented by either a parametric distance function or a set-theoretic DEA construction.²⁴ Here, too, all intended outputs and inputs are assumed

²⁴ See the citations in Footnote 10. The distance function (inaptly named since it does not satisfy the properties of a mathematical distance function) is a particular (homogeneous) representation of multiple-output technologies first formulated by Malmquist [1953] and Shephard [1953]. See Färe and Primont [1995] or Russell [1997] for a thorough treatment of this concept.

to satisfy standard disposability conditions, but two key assumptions are made regarding the unintended outputs. The first,

$$\langle x, y, z \rangle \in \tilde{T} \wedge \lambda \leq 0 \implies \langle x, \lambda y, \lambda z \rangle \in \tilde{T}, \quad (4.1)$$

is called “weak disposability”, a concept originally attributable to Shephard [1953, 1974]. The second,

$$\langle x, y, z \rangle \in \tilde{T} \wedge z = 0 \implies y = 0, \quad (4.2)$$

is called “null jointness”. Weak disposability and null-jointness imply that, (a) while pollution is not freely disposable, it is possible to jointly and proportionately decrease pollution and the intended outputs and (b) production of *any* positive level of intended output always results in positive amounts of the residual being generated. This literature is predicated on the belief that these two assumptions can capture the fact that, starting at any efficient point of the technology, it is not possible to decrease pollution without decreasing the production of the intended outputs, and hence that, together, they model the positive reduced-form correlation between pollution and other intended outputs.

As argued in Section 2, the treatment of a by-product as any other productive input is contrary to the intuition we have about the role by-products such as pollution play in intended production.²⁵ Two examples in Section 6 demonstrate that (i) the weak-disposability restriction on pollution-generating technologies does not preclude a *negative* correlation between intended and unintended outputs.²⁶ and (ii) it is possible to rationalize situations with abatement activities where no pollution is generated even with positive levels of intended output, a violation of null-jointness.²⁷

By explicitly modeling abatement activities as a part of both the process of intended production and the process of residual generation, the analysis in Section 3.2 above provides a theoretical foundation for the positive correlation between intended outputs and residuals that is assumed at the outset in the conventional literature. Moreover, the analysis in Section 3 demonstrates the existence of a much more fundamental cause of the positive correlation

²⁵ In the literature, the treatment of pollution as any other productive input is often justified by considering the amount of pollution generated as a proxy for the amount of the assimilative capacity of environmental resources such as air and water used to absorb the pollution generated. However, a clear distinction needs to be made between these environmental resources, which definitely serve as inputs into the production process, and pollution, which is an output of production. A given environmental resource like air can absorb different types of unintended outputs like CO₂, SO₂, *etc.*, and its assimilative capacity can be different for different pollutants. See Murty [2010] for this distinction and its general equilibrium consequences. In this paper we abstract from the possible use of environmental resources as inputs into the production process.

²⁶ A fact already noted in the literature cited in Footnote 10 above.

²⁷ This could be true, *e.g.*, in the presence of abatement activities where all wastes can be recycled or all wastes are biodegradable and hence can be eliminated.

than the existence of resource-costly abatement options, namely, the use of certain inputs like coal or the production of certain intended outputs that generate pollution. The fact that pollution is caused by such inputs or outputs implies that a positive correlation between intended and unintended outputs exists even in the absence of abatement options. In other words, the models of technology in the conventional literatures can be interpreted as ones where the production relation that characterizes the residual generation mechanism (or the set T_2) involves only the abatement activities of the firm and the unintended outputs. In the model of by-production developed in Section 3, however, the residual-generating mechanism is primarily a relation in nature between inputs and intended outputs that cause pollution, with the level of pollution generated being conditioned also by the abatement activities of the firm.²⁸

Further, we find that, in general, the assumptions made in the conventional literature about the disposability properties of the reduced-form technology are not borne out by \tilde{T} , which was derived in Section 3 from the intended production technology and the residual generation mechanism.²⁹ In particular, with respect to the reduced-form function \tilde{f} , note that

$$\tilde{f}_i(x^1, x^2, y, z) = f_a(x^1, x^2, y, y^a)h_i(x, y^{-j}, z) + f_i(x^1, x^2, y, y^a), \quad i = n_1 + 1, \dots, n. \quad (4.3)$$

Given (3.18) and the sign conventions in (3.14), the sign of \tilde{f}_i is ambiguous for a pollution-generating input i , contrary to the conventional literature, where it is signed as per a normal input.

5. Illustrative empirical application of by-production: measuring technical efficiency.

The foregoing analysis reveals that modeling the phenomenon of by-production requires more than one implicit production relation among inputs and outputs. One of these relations captures intended production activities of firms (that is, describes the set T_1), while the other captures the inevitability of residual generation when firms engage in intended production (that is, describes the set T_2). The former identifies an upper bound for the intended outputs of firms for every given level of inputs, while the latter identifies a lower bound for pollution generation given every level of intended outputs and inputs that are responsible for causing

²⁸ See the definition of T_2 in Section 3.2.

²⁹ This was apparent in Theorems 1 and 2 in Section 3, which showed that technologies that satisfy by-production do not satisfy free disposability in inputs that cause pollution and in the abatement output.

pollution. Combined with appropriate disposability assumptions, the implications of by-production are clear for econometric and DEA models of pollution-generating technologies.

The econometric approach must involve simultaneous estimation of two (or more) production relations that have the above features. In particular the production relation associated with intended production will be the upper frontier of T_1 and the production relation associated with residual generation will be the lower frontier of T_2 . These production relations should satisfy the trade-offs implied by (3.14).

We now demonstrate with the help of an empirical application how multiple production relations involved in the by-production approach can be captured in the activity-analysis-based DEA approach and how standard measures for technical efficiency can be derived by employing this approach. We consider a more general model than the one presented above, incorporating multiple pollution-generating inputs and multiple pollutants. Our data set, which is described below, does not have information regarding abatement efforts of firms. Moreover, pollution in our data set is caused by the use of certain inputs by firms.³⁰

We augment the notation in Section 2 as follows:

- (i) p decision making units, indexed by d .
- (ii) m intended outputs, indexed by j , with quantity vector $y \in \mathbf{R}_+^m$. The $p \times m$ matrix of observations on intended output quantities is denoted by Y .
- (iii) n inputs, indexed by i . The first n_1 are non-pollution-generating, while the remaining $n_2 = n - n_1$ are pollution generating. The quantity vector is $x = \langle x^1, x^2 \rangle \in \mathbf{R}_+^n$. The $p \times n$ matrix of observations on the input quantities is denoted by $X = \langle X^1, X^2 \rangle$.
- (iv) m' pollutants, indexed by k , with quantity vector $z \in \mathbf{R}_+^{m'}$. The $p \times m'$ matrix of observations on pollutants is denoted by Z .

We use annual data for 92 coal-fired electric power plants from 1985 to 1995.³¹ This data base includes observations for one intended output: net electricity generation, measured in kWh; two unintended outputs: sulfur dioxide (SO₂) and nitrogen oxide (NO_x); two non-polluting inputs: the capital stock and the number of employees; and three pollution-generating inputs: the heat content (in Btu) of coal, oil, and natural gas consumed at each power plant. Thus $p = 92$, $m = 1$, $m' = 2$, $n_1 = 2$, and $n_2 = 3$. In particular, for $d = 1, \dots, 92$, we denote the electricity output by y^d , sulfur dioxide output by z_1^d , nitrogen

³⁰ Extension to the case where some intended outputs also cause pollution is straightforward. In the next section we consider some numerical examples to illustrate the extension to the case with abatement efforts of firms.

³¹ We are indebted to Carl Pasurka for supplying us with these data. A detailed description of the data can be found in Pasurka [2006].

dioxide output by z_2^d , capital stock by x_1^d , labor by x_2^d , coal by x_3^d , oil by x_4^d , and natural gas by x_5^d . Thus, $x^{1^d} = \langle x_1^d, x_2^d \rangle$ and $x^{2^d} = \langle x_3^d, x_4^d, x_5^d \rangle$.

As discussed above, by-production implies that the pollution-generating technology is $T_1 \cap T_2$. We will denote the overall technology $T_1 \cap T_2$ that satisfies by-production by T_{BP} . Thus, a data set coming from pollution-generating units must simultaneously belong to both T_1 and T_2 .

5.1. Constructing T_1 .

We assume that T_1 satisfies free disposability of inputs, abatement output, and intended outputs (as defined in (2.2), (2.3), and (2.12)) and that it is closed, convex, and, for the purpose of this empirical exercise, satisfies constant returns to scale.³² In addition, T_1 satisfies the following assumption, which we call “independence of T_1 from z ” and which states that pollution does not directly affect production of intended outputs:³³

$$\langle x, y, z \rangle \in T_1 \implies \langle x, y, \bar{z} \rangle \in T_1 \quad \forall \bar{z} \in \mathbf{R}^{m'}. \quad (5.1)$$

The intended-output technology T_1 that satisfies these assumptions is obtained in a standard way using DEA techniques as follows:

$$T_1 = \left\{ \langle x, y, z \rangle \in \mathbf{R}^{n+m+m'} \mid \lambda X \leq x \wedge \lambda Y \geq y \text{ for some } \lambda \in \mathbf{R}_+^p \right\}. \quad (5.2)$$

5.1.2. Constructing T_2 .

We assume T_2 satisfies costly disposability of pollution, abatement output, and inputs that cause pollution (as defined in (3.15)). For the purpose of this empirical exercise, we also assume “constant returns to scale”: *e.g.*, doubling all pollution-causing inputs doubles the amount of all pollutants.³⁴ Also note that, since we have assumed that only x^2 affects residual generation, T_2 also satisfies “independence of T_2 from x^1 and y ”:

$$\langle x, y, z \rangle \in T_2 \implies \langle \bar{x}^1, x^2, \bar{y}, z \rangle \in T_2 \quad \forall \langle \bar{x}^1, \bar{y} \rangle \in \mathbf{R}^{n_1+m}. \quad (5.3)$$

The DEA version of T_2 , which satisfies these assumptions, is obtained as

$$T_2 = \left\{ \langle x^1, x^2, y, z \rangle \in \mathbf{R}^{n_1+n_2+m+m'} \mid \mu X^2 \geq x^2 \wedge \mu Z \leq z \text{ for some } \mu \in \mathbf{R}_+^p \right\}. \quad (5.4)$$

³² Modifications to non-decreasing or variable returns can be done in the usual way. (See, *e.g.*, Färe, Grosskopf, and Lovell [1995].) Assuming constant returns facilitates calculation of efficiency indexes using the method proposed by Färe, Grosskopf, and Zaim [2002].

³³ This assumption would have to be relaxed if, *e.g.*, the presence of pollution could adversely affect labor productivity in producing intended outputs.

³⁴ Footnote 32 still pertains.

The first inequality in (5.4) reflects costly disposability of inputs that cause pollution and the second reflects costly disposability of pollution. Since T_2 is independent of x^1 and y , no inequalities need to be specified for x^1 and y .

5.2. Constructing $T_{BP} = T_1 \cap T_2$.

The overall technology that exhibits by-production is the intersection of T_1 and T_2 :

$$T_{BP} = \left\{ \langle x^1, x^2, y, z \rangle \in \mathbf{R}^{N_1+N_2+M+M'} \mid \lambda[X^1 \ X^2] \leq \langle x^1, x^2 \rangle, \lambda Y \geq y, \right. \\ \left. \mu X^2 \geq x^2, \mu Z \leq z, \right. \\ \left. \text{for some } \langle \lambda, \mu \rangle \in \mathbf{R}_+^{2p} \right\}. \quad (5.5)$$

The above construction of T_{BP} using activity analysis involves two sets of production relations. These are reflected in the two different intensity vectors λ and μ , each of which is applied to the same data set.

In the analysis below, we compare the results from our by-production approach with those from the standard DEA approach based on the assumptions of weak disposability and null jointness.³⁵ To this end, we present below the comparable DEA formulation of the reduced-form technology denoted by \tilde{T}_{WD} used by this literature:

$$\tilde{T}_{WD} = \left\{ \langle x, y, z \rangle \in \mathbf{R}^{n+m+m'} \mid \lambda X \leq x \wedge \lambda Y \geq y \wedge \lambda Z = z \text{ for some } \lambda \in \mathbf{R}_+^p \right\}. \quad (5.6)$$

5.3. Measuring technical efficiency.

We calculate two types of efficiency indexes: the hyperbolic index employed in the original DEA study of Färe, Grosskopf, and Pasurka [1986] and the directional-distance-function index employed in more recent studies (*e.g.*, Färe, Grosskopf, Noh, and Weber [2005]). We calculate these indexes for both T_{BP} and \tilde{T}_{WD} .³⁶

For each technology $T = \tilde{T}_{WD}, T_{BP}$, for each year (1985–1995), and for each electricity plant ($d' = 1, \dots, 92$), the hyperbolic efficiency index is defined by³⁷

$$E_H(x^{d'}, y^{d'}, z^{d'}, T) = \min \left\{ \beta \mid \left\langle \beta x^{d'}, \frac{y^{d'}}{\beta}, \beta z^{d'} \right\rangle \in T \right\}. \quad (5.7)$$

³⁵ See, *e.g.*, Färe, Grosskopf, Lovell, and Pasurka [1989].

³⁶ The code for these programs and the data base can be accessed at <http://economics.ucr.edu/people/russell/index.html>

For greater details on the formulation, see the working paper version of this paper (Murty and Russell [2010]).

³⁷ The hyperbolic efficiency index was formulated by Färe, Grosskopf, and Lovell [1985, pp. 110–111].

For each technology $T = \tilde{T}_{WD}, T_{BP}$, for each year (1985–1995), and for each electricity plant ($d' = 1, \dots, 92$), the directional-distance-function efficiency index is defined by³⁸

$$E_{DD}(x^{d'}, y^{d'}, z^{d'}, T) = \min\{\beta \mid \langle x^{d'} - \beta g_x, y^{d'} + \beta g_y, z^{d'} - \beta g_z \rangle \in T\}, \quad (5.8)$$

where $g = \langle g_x, g_y, g_z \rangle \in \mathbf{R}^{n+m+m'}$ is the “direction vector.” E_H maps into the $(0,1]$ interval, while E_{DD} maps into \mathbf{R}_+ . For points on the frontier of T , $E_H(x, y, z, T) = 1$ and $E_{DD}(x, y, z, T) = 0$.

Simple comparisons of the hyperbolic index number results under the alternative assumptions about the technology are shown in Table 1.³⁹ The rank and product-moment correlation coefficients between the index number values range from .5 to .7 and from .6 to .8, respectively, indicating that the results are quite sensitive to the assumption made about the technology. The mean absolute difference between the index values is roughly in the .03 to .05 range (recall that the theoretical range of the index is 0 to 1), while the maximum difference is as high as .33 (and as low as .16) across the 11 years. As shown in Table 2, the correlation coefficients are even lower for the directional-distance-function index, ranging from .3 to .7.⁴⁰

Table 2 lists the number of efficient electricity plants—*i.e.*, the number of plants on the frontier of each of the technologies—in each year. The key message of this comparison is that the by-production technology tends to be much smaller than the reduced-form weak-disposability technology; that is the frontier of T_{BP} tends to be inside the frontier of \tilde{T}_{WD} . Moreover, while many firms are efficient relative to the \tilde{T}_{WD} technology but not the T_{BP} technology, *none* is efficient relative to the T_{BP} technology but not the \tilde{T}_{WD} technology. These comparisons, perhaps, are not surprising, since the former technology explicitly incorporates the constraints posed by the pollution-generating mechanism. (It should be noted, however, that the two technologies are not theoretically nested.) These comparison, in turn, suggest that the approach based on the reduced-form weak-disposability technology understates the amount of inefficiency (assuming of course that our by-production model is a better reflection of reality).

³⁸ The directional-distance-function index was adapted from the shortage function of Luenberger [1992] to the measurement of efficiency by Chambers, Chung, and Färe [1996] and Chung, Färe, and Grosskopf [1997]. For a comparison of the properties of these two efficiency indexes, among others, see Russell and Schworm [2010]. Following much of the literature, we choose the unit vector for g (*i.e.*, all elements of g are 1) in the calculation of the directional-distance-function indexes.

³⁹ The index values for all 92 plants over all 11 years can be found at <http://economics.ucr.edu/people/russell/index.html>

⁴⁰ Mean and maximal differences are not displayed for the directional-distance-function index, since its values are sensitive to the (arbitrary) size of the direction vector g . Alternatively, given the common choice of g as the unit vector, the values are sensitive to units of measurement.

6. By-production versus weak disposability: Comparisons of DEA formulations in the presence of abatement efforts.

In this subsection, we extend the DEA formulation of the technology satisfying by-production to include abatement efforts made by firms. Inclusion of abatement output facilitates a direct and meaningful comparison between the by-production and the weak-disposability approaches to modeling pollution generating technologies, as the latter approach, explains the positive correlation between intended outputs and pollution through abatement efforts of firms that are not modeled and hence considers only a reduced-form of the technology. A DEA version of the reduced-form by-production technology seen in (3.23) can be derived.

The DEA extension of the by-production technology to include abatement efforts of firms is simple. With respect to the intended technology T_1 , abatement is a standard output that satisfies standard output free disposability. The residual generating mechanism T_2 , on the other hand, satisfies costly disposability of abatement output. Thus,⁴¹

$$T_{BP} = \left\{ \langle x^1, x^2, y, y^a, z \rangle \in \mathbf{R}^{n_1+n_2+m+1+m'} \mid \begin{aligned} &\lambda[X^1 \ X^2] \leq \langle x^1, x^2 \rangle, \lambda Y \geq y, \lambda A \geq y^a, \\ &\mu X^2 \geq x^2, \mu A \leq y^a, \mu Z \leq z, \\ &\text{for some } \langle \lambda, \mu \rangle \in \mathbf{R}_+^{2D} \end{aligned} \right\}, \quad (6.1)$$

where A is the vector of abatement outputs for the p firms.⁴² The required reduced-form technology corresponding to T_{BP} is the projection of T_{BP} into the space of all goods other than the abatement output.

To compare the reduced form technologies derived from the two approaches, we hold input quantities fixed. Let $P_1(x)$, $P_2(x)$, $P_{BP}(x)$ and $\tilde{P}_{WD}(x)$ denote the restrictions of T_1 , T_2 , T_{BP} and \tilde{T}_{WP} , respectively, when inputs are held fixed at level x . Note that $P_{BP}(x)$ is defined in the space of all outputs: abatement, intended outputs, and unintended outputs.⁴³ The comparison of reduced-form technologies hence requires consideration of its projections into $\langle y, z \rangle$ space. Define

$$\begin{aligned} \hat{P}_1(x) &= \left\{ \langle y, y^a \rangle \in \mathbf{R}^{m+1} \mid \langle y, z, y^a \rangle \in P_1(x) \text{ for some } z \in \mathbf{R}^{m'} \right\} \text{ and} \\ \hat{P}_2(x) &= \left\{ \langle z, y^a \rangle \in \mathbf{R}^{m'+1} \mid \langle y, z, y^a \rangle \in P_2(x) \text{ for some } y \in \mathbf{R}^m \right\}. \end{aligned} \quad (6.2)$$

⁴¹ Again, this formulation maintains constant return to scale in both technologies, an assumption that can be changed to non-increasing or variable returns in the standard way.

⁴² Extension to include more than one abatement output is straightforward.

⁴³ For greater details, see the working paper version of this paper (Murty and Russell [2010]).

Then the projection of $P_{BP}(x)$ into the space of all intended and unintended outputs is given by

$$\hat{P}_{BP}(x) = \left\{ \langle y, z \rangle \in \mathbf{R}^{M+M'} \mid \langle y, y^a \rangle \in \hat{P}_1(x) \wedge \langle z, y^a \rangle \in \hat{P}_2(x) \text{ for some } y^a \in \mathbf{R}_+ \right\}. \quad (6.3)$$

The difference between the two DEA approaches is elucidated by a comparison of $\hat{P}_{BP}(x)$ and $\tilde{P}_{WD}(x)$.

We provide two numerical examples to compare the two reduced-form technologies derived by employing DEA methods. In both of the following examples, $n_2 = 1$, $n_1 = 0$, $m = m' = 1$, and $x = 1$.

Example 1: $p = 5$. The (artificial) data are as follows:

$$\begin{array}{cccc} x & y^a & y & z \\ 1 & 1 & 4 & 9 \\ 1 & 2 & 6 & 6 \\ 1 & 3 & 2 & 6 \\ 1 & 4 & 4 & 3 \\ 1 & 5 & 2 & 2 \end{array} \quad (6.4)$$

After plotting the data, we find that $\hat{P}_1(1)$ and $\hat{P}_2(1)$ can be represented functionally by piece-wise linear functions:

$$\begin{aligned} \hat{P}_1(1) &= \{ \langle y^a, y \rangle \in \mathbf{R}_+^2 \mid y \leq \rho^1(y^a) \} \text{ and} \\ \hat{P}_2(1) &= \{ \langle y^a, z \rangle \in \mathbf{R}_+^2 \mid z \geq \rho^2(y^a) \}, \end{aligned} \quad (6.5)$$

where

$$\begin{aligned} \rho^1(y^a) &= 6, & y^a &\in [0, 2] \\ &= 8 - y^a, & y^a &\in [2, 4] \\ &= 12 - 2y^a, & y^a &\in [4, 5] \end{aligned} \quad (6.6)$$

and

$$\begin{aligned} \rho^2(y^a) &= 12 - 3y^a, & y^a &\in [1, 2] \\ &= 9 - \frac{3}{2}y^a, & y^a &\in [2, 4] \\ &= 7 - y^a, & y^a &\in [4, 5] \\ &= 2, & y^a &\geq 5. \end{aligned} \quad (6.7)$$

The sets $\hat{P}_1(1)$ and $\hat{P}_2(1)$ are shown in Panels 1 and 2 of Figure 1. $\hat{P}_{BP}(1)$, shown in Panel 3 of Figure 1, is constructed as follows:

$$\hat{P}_{BP}(1) = \{ \langle z, y \rangle \in \mathbf{R}_+^2 \mid z \geq \rho^2(y^a) \wedge y \leq \rho^1(y^a) \wedge y^a \in [0, 5] \}. \quad (6.8)$$

Panel 4 of Figure 1 shows $\tilde{P}_{WD}(x)$. Note that the construction of $\hat{P}_{BP}(x)$ involves explicit reference to the abatement output: in particular, we have been able to parametrize the frontier of the set $\hat{P}_{BP}(x)$ in terms of the parameter y^a . No reference was made, however, to data on y^a in the construction of $\tilde{P}_{WD}(x)$.

As seen in Panel 4 of Figure 1, $\tilde{P}_{WD}(x)$ satisfies weak disposability and null jointness, but, the frontier has negatively sloped regions, indicating a negative correlation between intended and unintended outputs. The frontier of $\hat{P}(x)$, on the other hand, is everywhere non-negatively sloped.

Example 2: $p = 7$. The data are as follows:

$$\begin{array}{cccc}
 x & y^a & y & z \\
 1 & 0 & 8 & 9 \\
 1 & 1 & 7 & 6 \\
 1 & 2 & 6 & 8 \\
 1 & 3 & 6 & 3 \\
 1 & 4 & 1 & 2 \\
 1 & 5 & 4 & 0 \\
 1 & 6 & 2 & 0
 \end{array} \tag{6.9}$$

Proceeding as in Example 1, we can obtain the two piece-wise linear functions that represent $\hat{P}_1(1)$ and $\hat{P}_2(1)$ and the set $\hat{P}_{BP}(1)$. The sets $\hat{P}_1(1)$, $\hat{P}_2(1)$, and $\hat{P}_{BP}(1)$ are shown in Panels 1, 2, and 3 of Figure 2.⁴⁴ The frontier of $\hat{P}_{BP}(x)$ is non-negatively sloped. As seen in Panel 4 of Figure 2, $\tilde{P}_{WD}(x)$ satisfies weak disposability but violates null jointness. In our example, this is rationalized by the fact that abatement output of a firm can completely mitigate pollution even when it is producing positive amounts of the intended outputs.

7. Conclusions.

Pollution is an unintended output that cannot be freely disposed of. Underlying its production are a set of chemical and physical reactions that take place in nature when firms engage in the production of intended outputs. These natural reactions define nature's residual generation mechanism, which is a relation between the residuals generated and some inputs that are used or some intended outputs that are produced by the firm: hence, the inevitability of a certain minimal amount of pollution being generated when firms engage in intended production. We call this phenomenon by-production of pollution. The larger is the

⁴⁴ For detailed computations, see the working paper version of this paper (Murty and Russell [2010]).

scale of intended production, the more are the pollution-causing inputs being used or the more are the pollution causing intended outputs being produced, and hence, the more is the pollution generated. This provides the fundamental explanation for the positive correlation that is observed between intended production and residual generation.⁴⁵

Most of the existing literature attributes the observed positive correlation to abatement options available to firms. We show that abatement options of firms add to the phenomenon of by-production in strengthening the observed positive correlation between pollution generation and the production of intended outputs. The existing literature usually does not explicitly model abatement options of firms, but considers a reduced form of the firm's technology, which satisfies standard disposability assumptions with respect to all inputs and intended outputs.

We show that more than one implicit production relation is needed to capture all the technological trade-offs that are implied by the phenomenon of by-production. In particular, we show that by-production can be modeled by decomposing the technology into an intended-production technology and a residual-generation technology. The latter must exhibit costly disposal of pollution, as discussed in Murty [2010]. Abatement activities of firms can be added to the model as an additional factor in both the intended-production technology and the residual-generation technology. From this general model, we are able to derive a reduced form of the technology in the space of inputs, intended outputs, and unintended outputs that is in the spirit of that usually studied in the literature. Contrary to the usual literature, however, we find that the reduced-form technology violates standard disposability assumptions with respect to inputs and intended outputs that cause pollution. We derive implications from the phenomenon of by-production for the econometric and DEA specifications of pollution-generating technologies. We derive a DEA specification of technologies that satisfies by-production and employ it to measure technical efficiency of firms. Our illustrative empirical application indicates that the many results in the extensive literature using the methods first proposed by Färe, Grosskopf, and Pasurka [2006] (and surveyed in Zhou, Ang, and Poh [2008]) might be altered using the approach that we propose. This is a question to be answered by further research.

⁴⁵ Some of the literature has adopted physical science terminology to describe these relationships in terms of the "material balance" condition (see Ayres and Kneese [1969] and, more recently, Coelli, Lauwers, and van Huylenbroeck [2007]).

APPENDIX

Implicit function theorem: Let $f : \mathbf{R}_+^n \times \mathbf{R}_+^m \rightarrow \mathbf{R}^m$ be a continuously differentiable vector valued function with image $f(x, y) = z$, where $x \in \mathbf{R}_+^n$ and $y \in \mathbf{R}_+^m$. Let $\langle \bar{x}, \bar{y} \rangle \in \mathbf{R}_+^{n+m}$ be such that $f(\bar{x}, \bar{y}) = 0$ and the $m \times m$ matrix $\nabla_y f(\bar{x}, \bar{y})$ is full-row ranked (has a non-zero determinant). Then there exist neighborhoods U and V around \bar{x} and \bar{y} in \mathbf{R}_+^n and \mathbf{R}_+^m , respectively, and a continuously differentiable function $\Phi : U \rightarrow V$ with image $\Phi(x) = y$ such that, for all $x \in U$, we have $f(x, \Phi(x)) = 0$ and

$$\nabla_x \Phi(x) = - [\nabla_y f(x, \Phi(x))]^{-1} \nabla_x f(x, \Phi(x)).$$

Table 1
Weak disposability vs. by-production technologies:
Comparison of values of efficiency indexes.

Year	Spearman Correlation Coefficient (Hyp.)	Pearson Correlation Coefficient (Hyp.)	Mean Difference (Hyp.)	Maximum Difference (Hyp.)	Spearman Correlation Coefficient (D.D.F.)	Pearson Correlation Coefficient (D.D.F.)
1985	.67	.75	.025	.13	.50	.57
1986	.55	.72	.033	.15	.39	.46
1987	.58	.81	.023	.14	.55	.59
1988	.54	.73	.033	.13	.44	.62
1989	.55	.67	.026	.13	.50	.58
1990	.56	.69	.029	.14	.46	.52
1991	.47	.64	.031	.14	.40	.51
1992	.54	.69	.027	.13	.47	.60
1993	.50	.65	.030	.13	.35	.54
1994	.53	.62	.031	.14	.30	.53
1995	.67	.72	.051	.25	.64	.67

Notes: The columns marked Hyp. contain comparisons of the hyperbolic efficiency index values, while those marked D.D.F. contain comparisons of the directional-distance-function efficiency index values. The second and third columns show, respectively, the rank and product-moment correlation coefficients between the efficiency index values calculated using DEA formulations of \tilde{T}_{WD} and T_{BP} . The fourth and fifth columns contain, respectively, the mean absolute difference and the maximal difference between the the index values calculated under the alternative assumptions about the technology.

Table 2
Weak disposability vs. by-production technologies:
Numbers of efficient (frontier) plants.

Year	\tilde{T}_{WD}	T_{BP}	Year	\tilde{T}_{WD}	T_{BP}
1985	35	9	1991	40	8
1986	37	5	1992	37	10
1987	40	10	1993	45	7
1988	40	8	1994	43	6
1989	39	9	1995	35	9
1990	37	7			

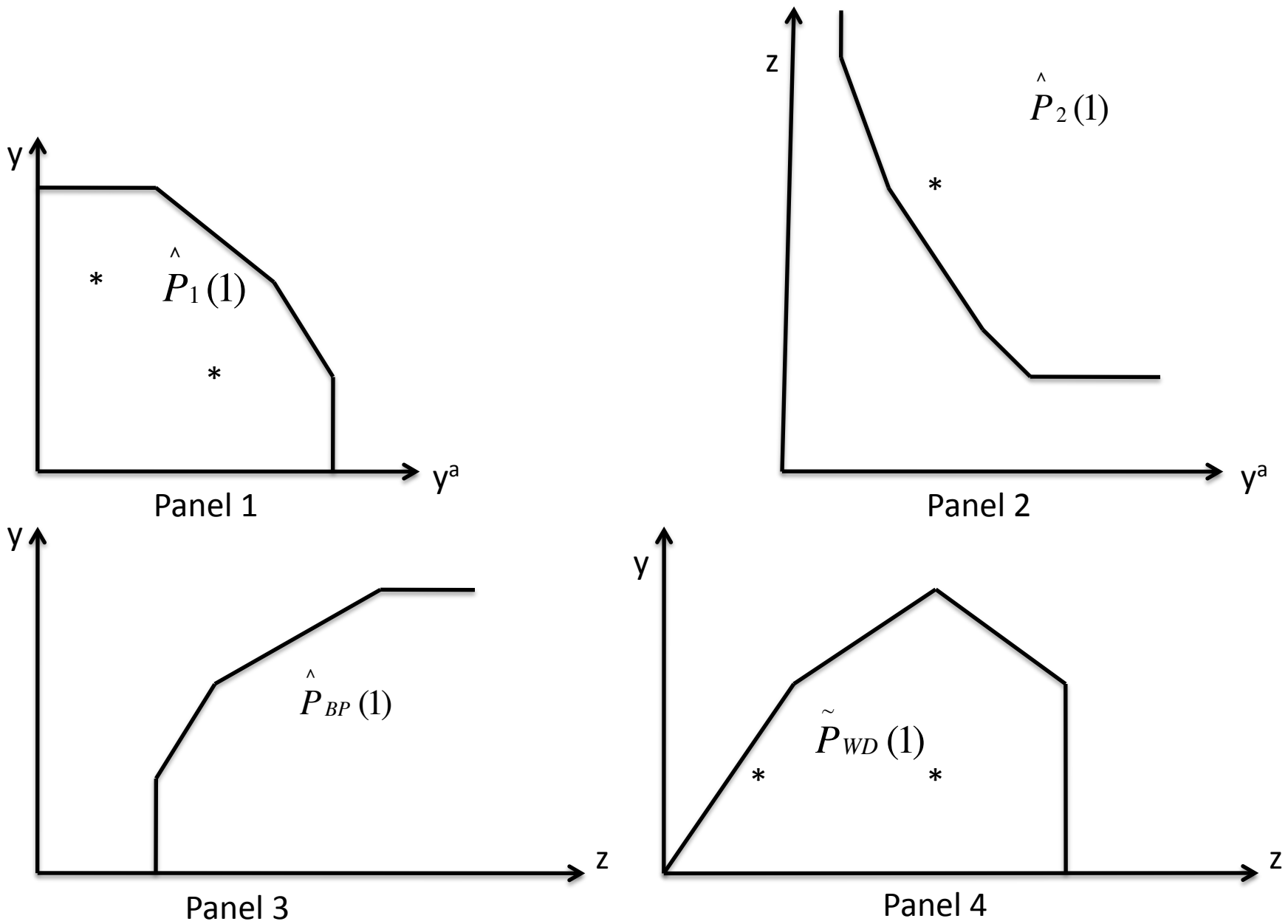


Figure 1

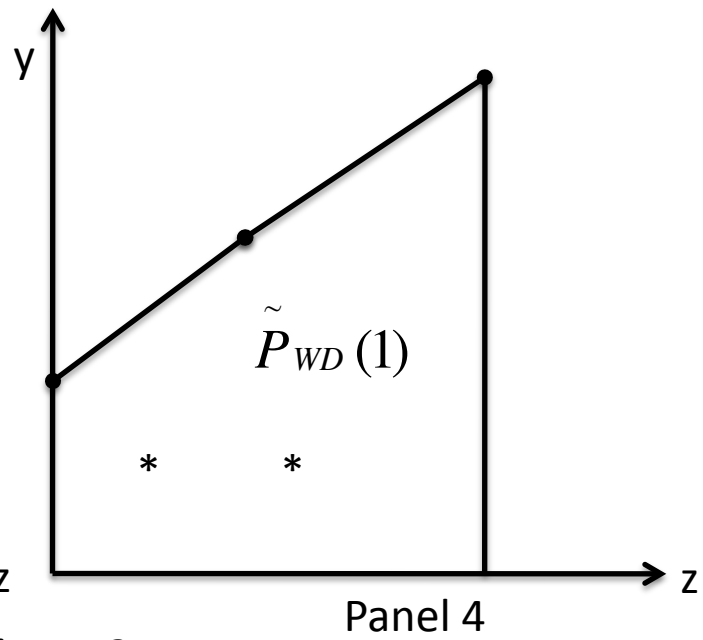
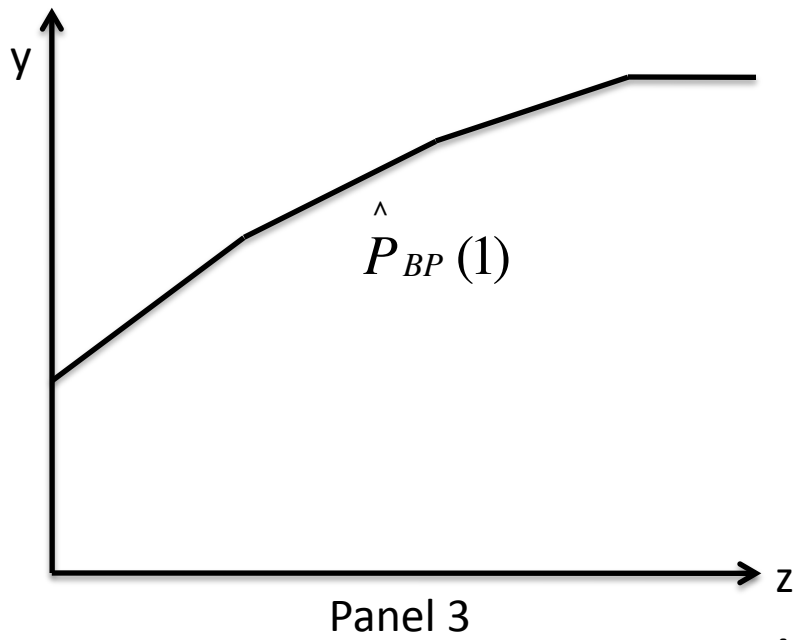
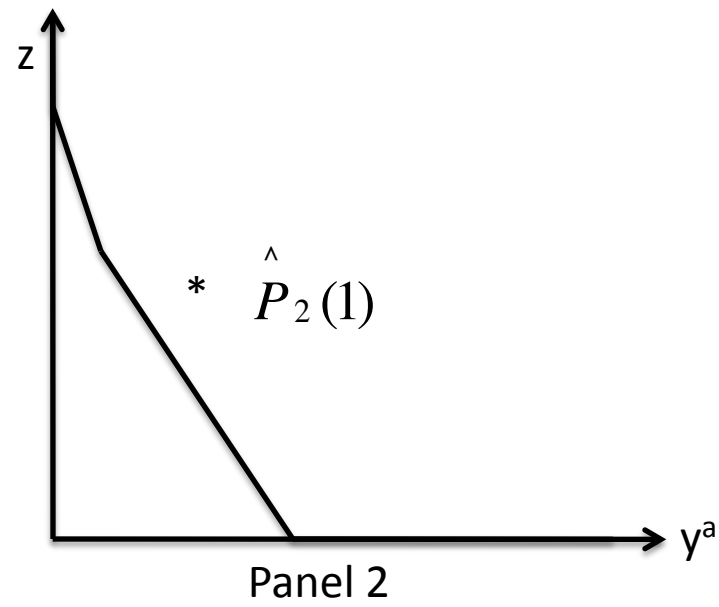
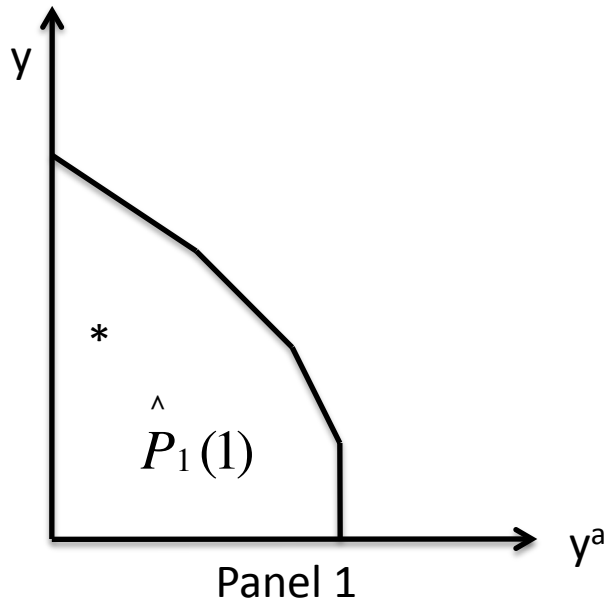


Figure 2

REFERENCES

- Aryes, R. U., and A. V. Kneese [1969], “Production, Consumption and Externalities,” *American Economic Review* 59: 282–297.
- Barbera, A. J., and V. McConnell [1990], “The Impact of Environmental Regulation on Industry Productivity: Direct and Indirect Effects,” *Journal of Environmental Economics and Management* 18: 50–65.
- Baumol, W. J., and W. E. Oates [1988], *The Theory of Environmental Policy*, Second Edition, Cambridge: Cambridge University Press.
- Boyd, G. A., and J. D. McClelland [1999], “The Impact of Environmental Constraints on Productivity Improvement in Integrated Paper Plants,” *Journal of Environmental Economics and Management* 38: 121–142.
- Chambers, R. G., Y. Chung, and R. Färe [1996], “Benefit and Distance Functions,” *Journal of Economic Theory* 70: 407–419.
- Chung, Y. H., R. Färe, and S. Grosskopf [1997], “Productivity and Undesirable Outputs: A Directional Distance Function Approach,” *Journal of Environmental Management*, 51: 229–240.
- Coelli, T., L. Lauwers, and G. van Huylenbroeck [2007], “Environmental Efficiency Measurement and the Materials Balance Condition,” *Journal of Productivity Analysis* 28: 3–12.
- Coggins, J. S., and J. R. Swinton [1996], “The Price of Pollution: A Dual Approach to Valuing SO₂ Allowances,” *Journal of Environmental Economics and Management* 30: 58–72.
- Cropper, M. L., and W. E. Oates [1992], “Environmental Economics: A Survey,” *Journal of Economic Literature* 30: 675–740.
- Färe, R., S. Grosskopf, and C. A. K. Lovell [1985], *The Measurement of Efficiency of Production*, Boston: Kluwer-Nijhoff.
- Färe, R., S. Grosskopf, and C. A. K. Lovell [1994], *Production Frontiers*, Cambridge: Cambridge University Press.
- Färe, R., S. Grosskopf, C. A. K. Lovell, and C. Pasurka [1989], “Multilateral Productivity Comparisons When Some Outputs are Undesirable: A Nonparametric Approach,” *The Review of Economics and Statistics* 71: 90–98.
- Färe, R., S. Grosskopf, C. A. K. Lovell, and S. Yaisawarng [1993], “Derivation of Shadow Prices for Undesirable Outputs: A Distance Function Approach,” *The Review of Economics and Statistics* 75: 374–380.
- Färe, R., S. Grosskopf, D. W. Noh, and W. Weber [2005], “Characteristics of a Polluting Technology: Theory and Practice,” *Journal of Econometrics* 126: 469–492.
- Färe, R., S. Grosskopf, and C. Pasurka [1986], “Effects of Relative Efficiency in Electric Power Generation Due to Environmental Controls,” *Resources and Energy* 8: 167–184.

- Färe, R., S. Grosskopf, and O. Zaim [2002], “Hyperbolic Efficiency and Return to the Dollar,” *European Journal of Operational Research* 136: 671–679.
- Färe, R., and D. Primont [1995], *Multi-Output Production and Duality: Theory and Applications*, Boston: Kluwer Academic Press, 1995.
- Førsund, F. [2009], “Good Modelling of Bad Outputs: Pollution and Multiple-Output Production,” *International Review of Environmental and Resource Economics* 3: 1–38.
- Fried, H., C. A. K. Lovell, and S. S. Schmidt [2008], *The Measurement of Productive Efficiency and Growth*, Cambridge: Cambridge University Press.
- Frisch, R. [1965], *Theory of Production*, Dordrecht: D. Reidel Publishing Company.
- Hailu A., and T. S. Veeman [2000], “Environmentally Sensitive Productive Analysis of the Canadian Pulp and Paper Industry, 1959–1994: An Input Distance Function Approach,” *Journal of Environmental Economics and Management* 40: 251–274.
- Luenberger, D. G. [1992], “New Optimality Principles for Economic Efficiency and Equilibrium,” *Journal of Optimization and Applications*, **75**: 221–264.
- Malmquist, S. [1953], “Index Numbers and Indifference Surfaces,” *Trabajos de Estadística*, 4: 209–242.
- Murty, M. N., and S. Kumar [2002], “Measuring Cost of Environmentally Sustainable Industrial Development in India: A Distance Function Approach,” *Environment and Development Economics* 7: 467–86.
- Murty, M. N., and S. Kumar [2003], “Win-Win Opportunities and Environmental Regulation: Testing of Porter Hypothesis for Indian Manufacturing Industries,” *Journal of Environmental Management* 67: 139–44.
- Murty, M. N., S. Kumar, and M. Paul [2006], “Environmental Regulation, Productive Efficiency, and Cost of Pollution Abatement: A Case Study of Sugar Industry in India,” *Journal of Environmental Management* 79: 1–9.
- Murty, S. [2010], “Externalities and Fundamental Nonconvexities: A Reconciliation of Approaches to General Equilibrium Externality Modeling and Implications for Decentralization,” *Journal of Economic Theory* 145: 331–53.
- Murty, S., and R. R. Russell [2010], “On Modeling Pollution-Generating Technologies,” The Warwick Economics Research Papers Series (TWERPS), No. 931.
- Pasurka, C. A. [2006], “Decomposing Electric Power Plant Emissions Within a Joint Production Framework,” *Energy Economics* 28: 26–43.
- Pittman, R. W. [1983], “Multilateral Productivity Comparisons with Undesirable Outputs,” *The Economic Journal* 93: 883–391.
- Russell, R. R. [1997], “Distance Functions in Consumer and Producer Theory,” Essay 1 in R. Färe, S. Grosskopf, and R. R. Russell, *Index Number Theory: Essays in Honor of Sten Malmquist*, Boston: Kluwer Academic Publishers, 7–90 .

- Russell, R. R., and W. Schworm [2010], “Properties of Inefficiency Indexes on \langle Input, Output \rangle Space,” Working Paper, University of California, Riverside, and University of New South Wales.
- Shephard, R. W. [1953], *Cost and Production Functions*, Princeton: Princeton University Press.
- Shephard, R. W. [1974], *Indirect Production Functions*, Meisenheim Am Glan: Verlag Anton Hain.
- Zhou P., B. W. Ang, and K. L. Poh [2008], “A Survey of Data Envelopment Analysis in Energy and Environmental Studies,” *European Journal of Operational Research* 189: 1–18.