Generalized Estimation of Productivity with Multiple Bad Outputs: the Importance of Materials Balance Constraints

Scott E. Atkinson
Department of Economics
University of Georgia
email: atknsn@uga.edu

Mike G. Tsionas
Department of Economics
Lancaster University Management School, LA1, 4YX, UK
and Montpellier Business School, 2300 Avenue des Moulins, 34080 Montpellier, France
email: m.tsionas@lancaster.ac.uk*

November 11, 2020

Abstract

Previous research has frequently estimated the directional technology distance function (DTDF) to more flexibly model multiple-input and multiple-output production, firm inefficiency, and productivity growth. For example, with firms such as electric utilities, one must model the production of good and bad outputs using good and bad inputs. Typically, all inputs and outputs are potentially endogenous. In previous work, we show how to identify a DTDF system using price equations based on profit maximization and compute optimal directions for measuring productivity change. However, this work has not imposed restrictions that limit substitution possibilities among inputs and outputs to a feasible set that is consistent with materials-balance constraints. Such constraints require that the weight of all inputs equals the weight of all outputs. The major innovation of this paper is that we include two types of functional relationships that impose the parametric analog of materials balance by modeling the generation of bad outputs and the use of bad inputs. The first requires that bad outputs are functionally related to good inputs and bad inputs. The second requires that bad inputs are functionally related to good inputs. We illustrate these methods using a balanced panel of 80 U.S. coal-fired electric generating plants from 1995-2005. Substantial differences are observed between the specification that includes the materials-balance constraints and the conventional approach that omits them, based on Bayes factors as well as measures of productivity and inefficiency. For many plants, improved management practices can reduce substantial inefficiencies in meeting emission constraints without reducing productivity growth.

JEL CODES: C11, C33, D24

KEY WORDS: Productivity and Competitiveness, directional technology distance function, productivity change with goods and bads, materials-balance equations

*Corresponding author.

1 Introduction

A considerable operations research literature employs a non-stochastic, linear programming methodology to calculate the extent to which decision-making units (DMUs) maximize desirable production from a set of inputs subject to resource constraints. The literature employs Data Envelopment Analysis (DEA) (as developed by Banker, Charnes, and Cooper, 1984; Charnes, Cooper, and Rhodes, 1978; Cooper, Seiford, and Zhu 2011; Zhu, 2014) to compute efficiency and productivity measures for evaluating DMUs. This process is complicated when bad or undesirable outputs are included. Clearly, they cannot be treated as coequal with good outputs.

The modeling of pollutants as bad outputs has been formalized in several ways. While a few researchers (e.g., Hailu and Veeman, 2001) treat bad outputs as inputs, the majority of traditional models incorporating environmental performance measures treat the emission of pollutants: (i) as undesirable outputs (e.g., Färe, Grosskopf, Lovell, and Pasurka, 1989; Chung, Färe, and Grosskopf, 1997; Färe and Grosskopf, 2004; Färe, Grosskopf, and Pasurka, 2007a; Färe, Grosskopf, Pasurka, 2007b; Wang, Wei, and Zhang, 2012; Wang, Wei, and Zhang, 2013); (ii) as multiplicative inverse outputs or as large constant–added additive inverse outputs (e.g., Scheel, 2001; Seiford and Zhu, 2002, 2005; Sahoo, Luptacik, and Mahlberg, 2011); (iii) using a separate technology for desirable and undesirable outputs (e.g., Murty, Russell, and Levkoff, 2012; Sueyoshi and Goto 2012); and (iv) using the materials balance principle (MBP) (e.g., Prasad and Calis, 1999; Coelli, Lauwers, and Van Huylenbroeck, 2007; Lauwers, 2009; Welch and Barnum, 2009; Hampf, 2014; Hampf and Rødseth, 2015; Rødseth, 2016; and Wang, Wei, and Huang, 2018).[1]

The MBP states that the total amount of mass (i.e., material or energy) in the inputs must equal the mass in desirable outputs plus the mass in the residuals, typically in the form of pollution (Ayres and Kneese, 1969; Coelli et al., 2007). In other words, the mass of the material inputs equals the mass of the desirable and undesirable outputs (Ebert and Welsch, 2007), thereby safisfying the first law of thermodynamics, which states that matter can be transformed but not created or destroyed. Hampf and Rødseth (2015) and Rødseth (2016) propose models that satisfy the MBP and, under certain conditions, both the first the second laws of thermodynamics. The second law states that: a) as energy is transformed, more and more of it is wasted; b) with a closed system, entropy (disorder or randomness) can only increase. When estimating a directional technology distance function (DTDF), Hampf and Rødseth (2015) generalize slightly the original concept of G-disposability and extend it to a formulation known as weak G-disposability. Under this condition, their model satisfies the first law of thermodynamics for given directions (typically denoted as a vector $\mathbf{g}$) of movement to the production frontier. Further, under weak G-disposability, the MB equation satisfies the second law of

[1] Dakpo, Jeanneaus, and Latruffe (2015) provide a useful summary of nonparametric models for works including MB constraints.

thermodynamics if one additionally requires that bad outputs must be strictly positive for a strictly positive amount of a bad input.

Wang et al. (2018) propose a DEA-based approach along with the MBP that accounts for the first law of thermodynamics to evaluate environmental and abatement efficiency simultaneously. The model incorporates weak G-disposability and also attempts to identify all possible adjustments to the mass of potentially polluting inputs bound up in inputs and outputs, as well as potential adjustments in the abatement of pollutants. They calculate the impact of each on efficiency measures. To accomplish this, they examine three different strategies to minimize pollution generation with three different degrees of freedom to adjust the polluting and non-polluting input mix.[2]

Wang, Zhou, Xie, and Zhang (2019) recognize that unobserved heterogeneity may cause inaccuracies in DEA calculations. They model the group heterogeneity of state-owned vs. locally-owned electric utilities by estimating separate DEA models for each and comparing these with the meta-frontier analysis which combines the two groups.

Many authors (e.g. Archibald, 1988; Hill, Vincent, and Chouinard, 1999; Rennings, 2000; Wossink, Lansink, and Struik, 2001) argue that production, pollution, and abatement have to be treated as non-separable. Therefore, pollution control can be achieved by changing production practices. Picazo-Tadeo and Prior (2009) show that, under weak disposability of pollution, firms may improve efficiency, leading to better economic performance without affecting environmental performance. De Koeijer, Wossink, Ittersum, and Struik (1999) and De Koeijer, Wossink, Struik, and Renkema (2002) show that using inputs more efficiently may improve both economic and environmental performance. To achieve this, it is necessary to drop the weak disposability assumption for pollution: environmental improvement does not always imply private costs since economic and environmental performances improve at the same time.[3]

Rødseth (2016) introduced a new efficiency measure and shows that the Coelli et al. (2007) model applied to pig farm production is likely to underestimate environmental efficiency when firms utilize pollution control. The reason is that Coelli et al. (2007), building on Lauwers (2009), only consider input minimization and substitution as possible compliance strategies, but ignore the potential additional costly resources needed to clean up pollutants (see also Pethig, 2006). Rødseth (2016) points out that one should credit the firm in a MB constraint for controlled emissions. For example, a pig farm may spend additional resources to clean up nitrogen emissions which would be ignored if the objective function was to minimize nitrogen inflows to the production process.[4] As an alternative to the Coelli

[2]They rightly mention that: "the comparison of the efficiency differences between modeling with MBP and without MBP shows that the MBP models provide a different but more accurate picture on the pattern of environmental and abatement efficiency, because the MBP models put more emphasis on reducing emissions and increasing abatement through both adjusting the sub-optimal polluting and non-polluting input mix and improving the quality of polluting input so as to reduce the polluting mass bound in the input" (p. 49).

[3]Dakpo, Jeanneaus, and Latruffe (2015) provide a summary of nonparametric models taking a MB approach.

[4]As he writes: "Farms A and B receive the same efficiency score (i.e., 1), although B's controlled emissions are lower

et al. (2007) approach, Hampf (2014) and Hampf and Rødseth (2015) introduce MB constraints that relate uncontrolled emissions to initially-generated pollution and controlled emissions.

In our study of coal-fired power plants, we adapt their MB constraint but modify their DEA approach to Stochastic Frontier Analysis (SFA). We assume the non-separability of production, pollution, and abatement. In our MB constraints, we employ measures of good and bad inputs to estimate initially-generated pollution, and measure controlled emissions with the plant's utilization of pre- and post-combustion control devices and plant-specific characteristics. The purpose of this approach is to minimize aggregation bias that may affect DEA estimates which use aggregate emission factors to relate bad inputs to initially-generated pollution. Unlike DEA studies, we stochastically impose the MBP in our estimation of efficiency and productivity, which allows us to calculate the efficiency with which plants achieve their MB constraints. This approach also allows us to model the observed and unobserved heterogeneity of DMUs, so that we can treat the potential endogeneity of inputs and outputs. Our results focus on productivity growth (and its components) and its relationship to measures of technical inefficiency (TI). We relate TI to technical efficiency (TE), using TI=1-TE. For a plant, TE is the measure, in percent terms relative to the frontier plant, of the maximum output using a given set of inputs for the production function or meet each MB constraint using a minimal set of inputs. We focus on TI since we directly measure it for each MB equation and the production function itself.[5] We impose the behavioral assumption of profit maximization on the estimation of the production frontier, including first-order conditions for profit maximization and the MB equations.

Since the SFA approach is stochastic, we obtain estimated standard errors and posterior densities using a Bayesian approach. This approach also allows us to present bivariate posterior densities that relate productivity growth (and its components) to many measures of firm behavior, including an overall measure of TI and a measure of TI for each MB constraint.

The coal-plants in our sample satisfy the first and second laws of themodynamics. They meet the first law of thermodynamics because we estimate our model with equations for the MB constraints. Our plants satisfy the second law of thermodynamics, since all of them use bad inputs to produce some bad outputs, so that MB governs the generation of the data.

than A's controlled emissions. Farm B is thus not rewarded for its efforts to control nitrogen emissions. The reason for this result is the Coelli et al. (2007) objective function which minimizes the nitrogen inflows to the production process. The additional capital and labor that farm B spends on manure transport relative to farm A are, in principle, seen as unproductive, since they do not contribute to improving the farm's environmental efficiency score. This suggests that a highly relevant environmental-economic trade-off (cf. Van Meensel, Lauwers, Van Huylenbroeck, and Van Passel (2010)) is neglected for the hypothetical sample, as well as for real-life samples where pollution control is a common compliance strategy (see Rødseth, 2016, p. 343)." Rødseth (2016) argues that the Coelli et al. (2007) approach only considers input minimization and substitution as possible compliance strategies, and ignores that additional resources may be used to clean up pollutants and, therefore, it does not reward pollution control.

[5] In terms of comparison between SFA and DEA, unsurprisingly we can attribute differences between efficiency scores principally to methodological issues. The measure of TI is higher under DEA because it is a non-stochastic approach that attributes any deviation from the frontier to TI, while SFA nets out random (idiosyncratic) error from TI. See Van Meensel et al. (2010), p. 1035

To our knowledge, our study is the first that combines MB constraints with a DTDF, which is a dual representation to (conformable with) profit maximization, with first-order conditions for profit maximization in an SFA context. Since we have a bad input (sulfur), we use separate MB constraints for the bad input and production of the bad outputs. The first MB constraint models the chemical bond between sulfur and one of the good inputs, carbon. We include in this constraint other good inputs in addition to carbon, since they allow the plant to burn coal with a higher sulfur or nitrogen content. We also include plant and firm dummies to capture differences in locale and plant efficiency. Secondly, we include MB constraints for uncontrolled emissions of each of three pollutants, which are sulfur dioxide (SO_2), carbon dioxide (CO_2), and nitrogen dioxide (NO_x). Uncontrolled emissions depend on initially-generated emissions and controlled emissions. Initially-generated emissions are a function of the bad input, good inputs, the pre-combustion technology, and the boiler firing type, while controlled emissions depend on good inputs and post-combustion control technologies.[6]

Since our analysis is in the SFA framework, several practical challenges have to be resolved before proceeding with formal estimation techniques. One problem is that quantities are endogenous, since they are arguably correlated with omitted variables. A second problem is that prices for inputs and outputs are missing at the plant level but are available at the firm level, so they are are potentially endogenous due to measurement error. We resolve these problems using the first-order conditions from profit maximization, firm-level prices, and instruments to adjust for measurement error and omitted variables. This approach also accounts for measurement error in the MB equations. We specify that the residuals of the MB equations are composed of one-sided and idiosyncratic terms, so that we can attribute inefficiencies in meeting the MB constraints to the former. Finally, we employ a Bayesian approach so that we can produce posterior densities for estimated coefficients and measures of efficiency and productivity growth, which yield information on the precision of the estimated statistics. To our knowledge, DEA has not been extended to allow for consideration of measurement error or statistical noise, although the bootstrap machinery can be potentially used to address such concerns. We find that including MB constraints makes a significant difference in estimated results for efficiency and productivity growth.

[6]This is similar to Førsund (2018) who introduces service inputs (such as labor, capital, and external services) that are not used up in the production process but improve the utilization of raw materials. Unlike his paper, we employ these variables as controls in our MB equations. Thus, our approach does not literally equate the weight of inputs to the weight of outputs.

2 The Directional Technology Distance Function and First-Order Conditions without Materials Balance

We assume a firm production technology in the absence of MB constraints that combines good inputs, $\mathbf{x} = (x_1, \ldots, x_N) \in R^N_+$, and bad inputs, $\tilde{\mathbf{x}} = (\tilde{x}_1, \ldots, \tilde{x}_J) \in R^J_+$, to produce good outputs, $\mathbf{y} = (y_1, \ldots, y_M) \in R^M_+$, and bad outputs, $\tilde{\mathbf{y}} = (\tilde{y}_1, \ldots, \tilde{y}_L) \in R^L_+$. The firm's production technology, $\mathcal{T}(t)$, at time t with vintage τ, can be written as

$$\mathcal{T}(t) = \{(\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y}, \tilde{\mathbf{y}}, \tau, t) : \mathbf{x}, \tilde{\mathbf{x}}\,\text{can produce}\,\mathbf{y}, \tilde{\mathbf{y}}\,\text{at time}\,t\;\text{with vintage}\,\tau\}, \tag{1}$$

where $t = 1, \ldots, T$.

Let $\boldsymbol{g} = (\boldsymbol{g}_x, \boldsymbol{g}_{\tilde{x}}, \boldsymbol{g}_y, \boldsymbol{g}_{\tilde{y}})$ be a direction vector. Some researchers assume that $(\boldsymbol{g}_x, \boldsymbol{g}_{\tilde{x}}, \boldsymbol{g}_y, \boldsymbol{g}_{\tilde{y}}) = (-1, -1, 1, -1)$, while other researchers do not scale inputs. Following Chambers (1998) and Chambers, Chung, and Färe (1998), we define the DTDF as

$$\begin{aligned} & \overrightarrow{D}_{\mathcal{T}}(\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y}, \tilde{\mathbf{y}}; \boldsymbol{g}, \tau, t) \\ = \; & \sup\{\beta : (\mathbf{x} + \beta \boldsymbol{g}_x, \tilde{\mathbf{x}} + \beta \boldsymbol{g}_{\tilde{x}}, \mathbf{y} + \beta \boldsymbol{g}_y, \tilde{\mathbf{y}} + \beta \boldsymbol{g}_{\tilde{y}}) \in \mathcal{T}\}. \end{aligned} \tag{2}$$

That is, the typical assumption is that the analyst measures the distance from the frontier using equal absolute values for directions which increase good outputs and reduce all other inputs and outputs.

We now specify the important properties of the DTDF that hold whether directions are assigned or estimated optimally. As shown in Hudgins and Primont (2007), for any values of the elements of $\boldsymbol{g}$, the following hold:

D1. Translation Property:

$$\begin{aligned} & \overrightarrow{D}_{\mathcal{T}}(\mathbf{x} + \alpha \boldsymbol{g}_x, \tilde{\mathbf{x}} + \alpha \boldsymbol{g}_{\tilde{x}}, \mathbf{y} + \alpha \boldsymbol{g}_y, \tilde{\mathbf{y}} + \alpha \boldsymbol{g}_{\tilde{y}}; \boldsymbol{g}_x, \boldsymbol{g}_{\tilde{x}}, \boldsymbol{g}_y, \boldsymbol{g}_{\tilde{y}}, \tau, t) \\ & \qquad = \overrightarrow{D}_{\mathcal{T}}(\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y}, \tilde{\mathbf{y}}; \boldsymbol{g}, \tau, t) - \alpha, \end{aligned} \tag{3}$$

D2. $\boldsymbol{g}$-Homogeneity of Degree Minus One:

$$\overrightarrow{D}_{\mathcal{T}}(\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y}, \tilde{\mathbf{y}}; \lambda \boldsymbol{g}_x, \lambda \boldsymbol{g}_{\tilde{x}}, \lambda \boldsymbol{g}_y, \lambda \boldsymbol{g}_{\tilde{y}}, \tau, t) = \lambda^{-1} \overrightarrow{D}_{\mathcal{T}}(\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y}, \tilde{\mathbf{y}}; \boldsymbol{g}, \tau, t), \;\; \lambda > 0, \tag{4}$$

D3. Concavity:

$$\overrightarrow{D}_{\mathcal{T}}(\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y}, \tilde{\mathbf{y}}; \boldsymbol{g}, \tau, t) \text{ is concave in } (\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y}, \tilde{\mathbf{y}}; \boldsymbol{g}, \tau, t), \tag{5}$$

D4. Non-negativity:

$$\overrightarrow{D}_{\mathcal{T}}(\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y}, \tilde{\mathbf{y}}; \boldsymbol{g}, \tau, t) \geq 0, \qquad (\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y}, \tilde{\mathbf{y}}; \boldsymbol{g}, \tau, t) \in \mathcal{T}. \tag{6}$$

Property *D1* is imposed parametrically, *D2* follows directly from *D1*, while *D4* can easily be imposed after estimation via normalization. We must test for *D3*. See Atkinson, Primont, Tsionas (2018) (APT) for details

We make the typical assumption of strong (S) disposability of good inputs and outputs:

D5–S. Good Input Monotonicity: Assuming that good inputs are strongly disposable, Appendix A.1 of APT shows that this implies:

$$\partial\overrightarrow{D}_{\mathcal{T}}(\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y}, \tilde{\mathbf{y}}; \boldsymbol{g}, \tau, t)/\partial x_n \geqq 0, n = 1, \ldots, N, \tag{7}$$

D6–S. Good Output Monotonicity: If we assume that good outputs are strongly disposable, following the proof of D5-S, in Appendix A.2 of APT, this assumption implies:

$$\partial\overrightarrow{D}_{\mathcal{T}}(\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y}, \tilde{\mathbf{y}}; \boldsymbol{g}, \tau, t)/\partial y_m \leq 0, m = 1, \ldots, M. \tag{8}$$

Temporarily suppressing all the arguments of $\overrightarrow{D}_{\mathcal{T}}(\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y}, \tilde{\mathbf{y}})$, we following Chambers (1998) and assume that a firm or plant maximizes profits, π, by choosing values of $(\mathbf{x}, \mathbf{y}, \tilde{\mathbf{x}}, \tilde{\mathbf{y}})$ to solve

$$\sup\left\{\mathbf{p}_y(\mathbf{y} + \overrightarrow{D}_{\mathcal{T}}\boldsymbol{g}_y) - \mathbf{p}_{\tilde{y}}(\tilde{\mathbf{y}} + \overrightarrow{D}_{\mathcal{T}}\boldsymbol{g}_{\tilde{y}}) - \mathbf{p}_x(\mathbf{x} + \overrightarrow{D}_{\mathcal{T}}\boldsymbol{g}_x) - \mathbf{p}_{\tilde{x}}(\tilde{\mathbf{x}} + \overrightarrow{D}_{\mathcal{T}}\boldsymbol{g}_{\tilde{x}})\right\}, \tag{9}$$

where $\mathbf{p}_y \geq 0, \mathbf{p}_{\tilde{y}} \geq 0, \mathbf{p}_x \geq 0$, and $\mathbf{p}_{\tilde{x}} \leq 0$ are price vectors, and the econometrician either predetermines or estimates $\boldsymbol{g}$. Further, we define $\mathbf{p} = (\mathbf{p}_y, \mathbf{p}_{\tilde{y}}, \mathbf{p}_x, \mathbf{p}_{\tilde{x}})$.

Typically, the analyst estimates the DTDF without behavioral assumptions such as PM, subject to a set of *a priori* fixed directions. However, in this paper we assume that the firm chooses $(\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y}, \tilde{\mathbf{y}})$ to maximize profits. This requires that we estimate the DTDF jointly with the first-order conditions for PM. Calculated directions, consistent with these conditions, are termed optimal-PM directions.[7] The first-order conditions (assuming positive values for all inputs and outputs) are:

$$p_n/\varrho(\mathbf{p}, \boldsymbol{g}) = \partial\overrightarrow{D}_{\mathcal{T}}(\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y}, \tilde{\mathbf{y}}; \boldsymbol{g}, \tau, t)/\partial x_n, n = 1, \ldots, N, \tag{10}$$

$$p_m/\varrho(\mathbf{p}, \boldsymbol{g}) = -\partial\overrightarrow{D}_{\mathcal{T}}(\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y}, \tilde{\mathbf{y}}; \boldsymbol{g}, \tau, t)/\partial y_m, m = 1, \ldots, M, \tag{11}$$

$$p_j/\varrho(\mathbf{p}, \boldsymbol{g}) = \partial\overrightarrow{D}_{\mathcal{T}}(\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y}, \tilde{\mathbf{y}}; \boldsymbol{g}, \tau, t)/\partial \tilde{x}_j, j = 1, \ldots, J, \tag{12}$$

$$p_l/\varrho(\mathbf{p}, \boldsymbol{g}) = \partial\overrightarrow{D}_{\mathcal{T}}(\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y}, \tilde{\mathbf{y}}; \boldsymbol{g}, \tau, t)/\partial \tilde{y}_l, l = 1, \ldots, L, \tag{13}$$

[7] The CM model and optimal-CM directions are obtained by using only the first-order conditions for input prices and assuming that outputs are given.

where

$$\varrho(\mathbf{p}, \boldsymbol{g}) = [\sum_m p_m g_m - \sum_n p_n g_n - \sum_j p_j g_j - \sum_l p_l g_l]. \tag{14}$$

We assume that $\varrho(\mathbf{p}, \boldsymbol{g}) > 0$, where $\varrho(\mathbf{p}, \boldsymbol{g})$ is the optimal value of the Lagrangian multiplier, which is the change in profits due to a small improvement in the production technology.

We also require assumptions about the prices of inputs and outputs. All prices are assumed to be non-negative except for the prices of bad inputs, which are non-positive. Bad inputs have non-positive prices since the firm will not pay to use such inputs, which are organically bound to good inputs. The prices of bad outputs are non-negative, since the firm must pay a fine or buy emission permits for their additional production. The prices of good outputs and the prices of good inputs are non-negative by definition.

The standard assumption of papers that attempt to estimate a DTDF without MB constraints is weak (W) disposability of bad inputs relative to good inputs and that $\mathbf{p}_{\tilde{x}} \leq 0$ in (12) obtains locally:

D7–W. Bad Input Monotonicity:

$$\partial \overrightarrow{D}_{\mathcal{T}}(\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y}, \tilde{\mathbf{y}}; \boldsymbol{g}, \tau, t)/\partial \tilde{x}_j \leq 0, j = 1, \ldots, J. \tag{15}$$

The standard assumption in papers that attempt to estimate a DTDF without MB constraints is weak disposability of bad outputs relative to good outputs which is sufficient for $\mathbf{p}_{\tilde{y}} \geq 0$ in (13) to obtain locally:

D8–W. Bad Output Monotonicity:

$$\partial \overrightarrow{D}_{\mathcal{T}}(\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y}, \tilde{\mathbf{y}}; \boldsymbol{g}, \tau, t)/\partial \tilde{y}_l \geq 0, l = 1, \ldots, L. \tag{16}$$

Following this approach we would maintain PM and use assumptions about the prices of bad inputs and bad outputs to locally restrict the range of the partial derivatives in D7–W and D8–W. The CM model utilizes only the monotonicity conditions for good and bad inputs, equations (12) and (13), which are the same as with the PM model.

However, Pethig (2006) and Murty, Russell, and Levkoff (2012) show that estimating the DTDF without the MB conditions and assuming weak disposability of good and bad outputs will fail to model abatement activity explicitly and will violate the MB principle.[8] They show that one can overcome this by using MB constraints to model gross residuals and abatement activities which control pollutants while modeling the intended production technology. Since we incorporate MB constraints in the remainder of this paper, we drop the assumption of the weak disposability of good and bad outputs given in D8-W.[9]

[8] Førsund (2018) makes a similar argument.

[9] Murty, Russell, and Levkoff (2012) specify a production technology mapping inputs into outputs that satisfies strong

We now turn to our parametric specification of the production technology, the restrictions that impose the translation property, and the parametric derivation of the first-order conditions for PM and CM. It is preferable at this point to assume that directions have unit length, viz., $||\mathbf{g}||_2 = 1$.

The DTDF is specified as a quadratic function for plant i in time period t:

$$\begin{aligned} 0 \; = \; & \sum_{w=1}^{W} \gamma_w z_{w,it} + \sum_{w=1}^{W} \sum_{w'=1}^{W} \gamma_{ww'} z_{w,it} z_{w',it} + \sum_{w=1}^{W} \gamma_{wt} z_{w,it}\, t + \sum_{w=1}^{W} \gamma_{w\tau} z_{w,it} \tau_{it} \\ + \; & \sum_{t=1}^{T} \gamma_t d_t + \sum_{i=1}^{N} \gamma_i d_i + \gamma_\tau \tau_{it} + \epsilon_{it}, \end{aligned} \tag{17}$$

where

$$\epsilon_{it} = \varepsilon_{it} - u_{it}, \tag{18}$$

so that ϵ_{it} is an additive error with a one-sided component, $u_{it} \geq 0$, and a standard noise component, ε_{it}, with zero mean, reflecting errors in optimization due to random events beyond the control of the firm. We specify that d_t is a year dummy, d_i is a plant-level dummy, and τ_{it} is vintage. The vector $\mathbf{z} = (\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y}, \tilde{\mathbf{y}})$ and $W = N + J + M + L$ for N good inputs, J bad inputs, M good outputs, and L bad outputs. Using a likelihood ratio test, we accept the null hypothesis that all coefficients for interaction terms of $z_{w,it}$ with t and τ_{it} are zero at the .05 level using a two-tailed test and impose this restriction henceforth.

Following APT, to guarantee that the translation property in (3) holds for the DTDF, we impose the following restrictions on the $\boldsymbol{\gamma}$ and $\boldsymbol{g}$ parameters of the DTDF and the associated share equations:

$$\begin{gathered} \sum_{w=1}^{W} \gamma_w g_w = -1, \;\; \sum_{w=1}^{W} \gamma_{wn'} g_w = 0, \;\; \forall\, n', \;\; \sum_{w=1}^{W} \gamma_{wm'} g_w = 0, \;\; \forall\, m', \\ \sum_{w=1}^{W} \gamma_{wj'} g_w = 0, \;\; \forall\, j', \;\; \sum_{w=1}^{W} \gamma_{wl'} g_w = 0, \;\; \forall\, l'. \end{gathered} \tag{19}$$

We can rewrite the first-order PM conditions from (10)-(13) in terms of the parameters of the quadratic DTDF in (17) for each good input price equation as:

$$p_{n,it}/\varrho = \gamma_n + \sum_{w=1}^{W} \gamma_{nw} (z_{w,it}) + v_{n,it}, \tag{20}$$

disposability properties, and by assumption is independent of the level of pollution. In particular, they reject the weak disposability of good and bad outputs due to the use of abatement activities. Importantly, in our paper, we do not assume the independence of the good and bad outputs.

for each good output price equation as:

$$p_{m,it}/\varrho = -\left[\gamma_m + \sum_{w=1}^{W} \gamma_{mw}(z_{w,it})\right] + v_{m,it}, \tag{21}$$

for each bad input price equation as:

$$p_{j,it}/\varrho = \gamma_j + \sum_{w=1}^{W} \gamma_{jw}(z_{w,it}) + v_{j,it}, \tag{22}$$

and for each bad output price equation as:

$$p_{l,it}/\varrho = \gamma_l + \sum_{w=1}^{W} \gamma_{lw}(z_{w,it}) + v_{l,it}, \tag{23}$$

where we impose symmetry on the parameters and where the $v_{w,it},\ w = n, m, j, l$ have zero mean. The DTDF *system* is comprised of the DTDF function in (17) and the price equations in (20)-(23), where we have imposed the parametric restrictions guaranteeing the translation property. This system was developed and estimated in APT.

3 Alternative Optimization Models

3.1 The New Model: Materials-Balance Constraints Plus Latent Prices

According to the technology set in (2), coal-fired power plants generate good and bad outputs using good and bad inputs. Their good output is electricity, their bad outputs are three pollutants (SO_2, NO_x, and CO_2), their good inputs are capital, labor, and heat input from coal, and their bad input is sulfur. The plants residual generation mechanism restricts the standard technology set in (2), since plants employ pre- and post-combustion pollution control devices, $\tilde{x}_a$, to reduce the generation of pollutants or their release into the environment. To translate the DEA model of Murty et al. (2012) to a stochastic framework we begin by formulating a scalar version of their residual-generation mechanism, by incorporating $\tilde{x}_a$:

$$F(x, \tilde{x}, y, \tilde{y}, \tilde{x}_a) = 0, \tag{24}$$

where we temporarily collapse our three good inputs into one generic good input, x, all pollution control devices into a single generic pollution control variable, $\tilde{x}_a$, and all pollutants into a single pollutant, $\tilde{y}$. From this we derive our two MB constraints which model the residual-generation mechanisms of the plant and limit the substitution possibilities implied by the transformation function in (2). Assuming that $\frac{\partial F}{\partial \tilde{y}} \geq 0$, since increasing the bad output will increase the distance to the frontier, following

Murty et al. (2012) we apply the implicit function theorem to (24) in a local neighborhood around $< x, \tilde{x}, y, \tilde{y}, \tilde{x}_a >$ to obtain the first MB production function:

$$\tilde{y} = g(\tilde{x}, x, y, \tilde{x}_a). \tag{25}$$

The second MB production function says that the good output is a function of only the good input (but not the bad output):

$$y = h(x). \tag{26}$$

Combining (25) and (26) yields our first MB constraint, where the bad output is a function of the bad input, the good input, and the pollution control device:

$$\tilde{y} = g^1_{MB}(\tilde{x}, x, \tilde{x}_a). \tag{27}$$

We assume that

$$\frac{\partial \tilde{y}}{\partial \tilde{x}} \geq 0, \tag{28}$$

since the change in the bad output should be non-negative as the bad input increases.[10]

$$\frac{\partial \tilde{y}}{\partial x} \geq 0, \tag{29}$$

since the change in bad output should be non-negative as the good input increases, and

$$\frac{\partial \tilde{y}}{\partial \tilde{x}_a} \leq 0, \tag{30}$$

since the change in the bad output should be non-positive as pollution control increases.

The constraint in (27) models, for example, a plant that burns a higher sulfur coal but employs flue gas desulfurization (FGD) devices, also called scrubbers, to reduce SO_2 emissions. Before estimation, we add plant/firm dummies to this MB equation to account for differences among plant and firm characteristics which themselves are time-invariant.

The second MB constraint relates the consumption of the bad input to the consumption of the good input because they are chemically bonded for all varieties of coal:

$$\tilde{x} = g^2_{MB}(x). \tag{31}$$

Before estimation, we add plant/firm dummies to incorporate differences among firms in terms of plant

[10] This occurs primarily for SO_2 and to a lesser degree for CO_2 and NO_x. The emission factors of all pollutants (the average emissions per million British Thermal Unit (Btu)) increase with the sulfur content of the coal burned.

management and location.

The two MB constraints in (27) and (31) generalize those previously employed in the literature. For Coelli et al. (2007) the MB constraint is

$$b = \sum_n u_n x_n - \sum_m v_m y_m, \tag{32}$$

where b is uncontrolled emissions, u_n translates input x_n into generation of the pollutant and v_m translates y_m into the amount of b bound up in y_m. Their application is to pig farms, where pigs contain a substantial portion of the bad nitrogen output.

Rødseth (2016), who also examines pig farms, recognizes that Coelli et al. (2007) do not reward pollution control. His revised MB constraint does so by requiring that the sum of uncontrolled and controlled emissions equals the generated pollution minus the amount of the pollutant bound up in the output:

$$b = \sum_n u_n x_n - \sum_m v_m y_m - a, \tag{33}$$

where a is the controlled emissions, i.e., emissions reduced by abatement devices.

Hampf (2014) and Hampf and Rødseth (2015) employ this MB constraint for the release of SO_2 and carbon dioxide CO_2, respectively, by coal-fired power plants. Both papers recognize that since no pollutants are bound up with electricity, $\sum_m v_m y_m = 0$. To estimate $\sum_n u_n x_n$, both papers utilize emissions factors which give the initial generation of SO_2 per unit of sulfur or the initial generation of CO_2 per unit of coal consumed, as reported in the 2009 DOE Electric Power Annual Appendix A1. This Appendix provides these emission factors by fuel and boiler firing type. Multiplying these emission factors by the amount of sulfur or coal consumed, respectively, for each plant yields estimates of total initial generation of each pollutant. Finally, by subtracting b from this estimate, they obtain a.

As Hampf (2014) recognizes, these aggregate emission factors are valid only if plants do not use different pre-combustion controls. However, many of our coal-fired plants make use of pre-combustion controls for SO_2, such as various types of sorbent injection. In addition, for these factors to be valid across plants, they must employ uniform management skills and operate a given boiler-firing technology with equal efficiency. Therefore, the use of aggregate SO_2 emission factors may create substantial aggregation bias. We believe that this potential bias is even greater for the nitrogen dioxide NO_x MB equation. Over time, an increasing number of our plants adopt different sets of pre-combustion and post-combustion NO_x controls in different years. Hence, we avoid using DOE Appendix A1 aggregate emission factors to estimate initially-generated pollution.

Instead, we take a stochastic approach that should substantially reduce any aggregation bias in the MB constraints for SO_2 and NO_x. Utilizing (33) for SO_2, we measure initially-generated pollution for SO_2 by including all plant-level good inputs, the bad input sulfur, plus plant and firm dummies. These dummies capture a variety of different boiler-firing configurations and different sorbent injection pre-combustion control methods that do not change over time for any plant and hence are time-invariant. We also include measures of controlled emissions, a. Since plants use scrubbers for additional SO_2 control, we include a variable measuring the magnitude and date of their installation. In our NO_x MB constraint, we include pre-combustion controls along with good inputs and the bad input to model initially-generated NO_x. We include the major post-combustion control technology to model controlled emissions. Finally, we obtain estimates of uncontrolled emissions, b, from EPA as explained in the Data section.

First, in greater detail with regard to the right-hand-side of our NO_x MB constraint, we include good inputs (the Btu content of coal, labor, and capital) to control for the amount of good output generated and hence initially-generated NO_x.[11] We do not include the nitrogen content of coal, since it is not reported to the DOE or EPA. However, the amount of NO_x produced from a pound of nitrogen depends almost totally on the pre-combustion technology of the plant. Hence, we include measures of pre-combustion control methods. Nearly all of our plants ultimately adopt either or both of two pre-combustion methods to reduce NO_x emissions: a low-NO_x boiler (LNOX) (which reduces the combustion temperature and hence the emissions of NO_x) or Overfire Air (OA) (which utilizes a special port designed to ensure advantageous air and fuel mixing, again to reduce NO_x). We include the annual MW capacity associated with each plant's boilers that utilize LNOX or OA. Many of our plants also employed post-combustion control devices termed Selective Catalytic Reduction (SCR) to remove NO_x from the exhaust gas. We also include the annual MW capacity for generators where SCRs reduce emissions from associated boilers. All of these variables are time-varying. In fact plants rarely used any of these technologies at the beginning of our sample, but by the end of our sample, nearly all plants had adopted at least one pre-combustion technology and the SCR technology. Finally, by including plant and firm dummies, we control for time-invariant differences in individual plant technologies and plant/firm management skills for NO_x abatement.

Next, with regard to the SO_2 MB constraint, we assume that the amount of SO_2 initially generated from the burning of coal depends on the good inputs (the Btu content of coal, labor, and capital which produce the good output required), the total sulfur content of coal, the pre-combustion use of sorbents, the firing type of the boiler burning the coal, and post-combustion controls. The most common types are cyclone, front, opposed, and tangential. The non-uniform use of sorbents, the exact sorbent used, and the boiler firing types remain constant over time for each plant, so that time-invariant differences among plants can be captured by plant dummies, which we include. The amount of SO_2 abated also depends of the use of FGD devices to treat post-combustion SO_2 emissions. Only 18 of our plants

[11] We note that output of electricity is a function of these same inputs, so that output does not appear in the MB constraint. Also, we include the sulfur content although its impact on NO_x production should be less important than nitrogen content of coal.

employ FGD devices throughout the entire sample period. We include in the MB equation the MW capacity of generators whose associated emissions are scrubbed using FGD devices for each plant in each year.[12] To control for any time-invariant differences in plant technologies and plant/firm management abilities for SO_2 removal, we include plant and firm dummies.[13]

Considering the CO_2 MB constraint, the amount of this pollutant released when coal is burned is fundamentally a function of the carbon content of the fuel. Unfortunately, the carbon content of coal is not reported at the plant level in DOE data files. However, Btu content is reported (which is our measure of coal consumed) and this is the best direct measure of the carbon content of coal.[14] During our sample period, no plants controlled CO_2 emissions. Thus, our MB equation for CO_2 depends on good inputs (the Btu content of coal, labor, and capital) plus plant and firm dummies. These dummies capture any time-invariant differences in individual plant boiler technologies and plant/firm management skills.

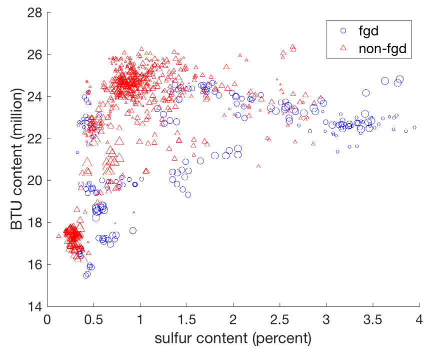
Implicit in the specification of the directional distance function in (9) and (10)–(13) is the differentiability of the directional distance function with respect to its arguments. In particular, this applies to the good inputs, $\mathbf{x}$, and the bad inputs, $\tilde{\mathbf{x}}$. While each mine typically sells one type of coal of a given approximate Btu and sulfur content (measured as a percent of coal), each plant purchases coal from a portfolio of many mines, each with a different combination of Btu and sulfur content. This is confirmed by examining the shipment-specific Energy Information Agency (EIA) Form 423 files for all plants and is illustrated in Figure 1 where we plot the Btu content per ton of coal against sulfur content for all shipments of coal from mines to power plants in our sample, obtained from this form. We find the greatest diversity in the combination of Btu per ton and sulfur content for the FGD plants, which comprise approximately 34% of total megawatt hour (MWh) generation.

[12] We do not include sorbent consumption or the hours that a scrubber is operated, since coal-fired power plants are base-load and for efficiency reasons plants and scrubbers are always operated at a near-constant rate. Thus, these variable inputs are treated as constant over time for a given plant and are swept up in plant dummies.

[13] Other characteristics such as moisture and ash content affect heat output but have little effect on the conversion of sulfur to SO_2, which occurs when sulfur and oxygen molecules are bonded in the combustion process.

[14] The discrepancy between the carbon content of coal and its Btu content during coal combustion is minor and would be due to the hydrogen content of the coal and, to a lesser extent, its sulfur content, which we also include. Hydrogen is not reported as a characteristic of delivered coal to plants in DOE data files. However, it is 5% or less of coal and little of this will be converted to heat since hydrogen combines with oxygen to produce water vapor. Sulfur generates little heat since it is a small component of coal by percent and has a very low heating value (about 3,000 Btu/ton).

Figure 1: Sulfur and Btu Content



Given a continuum of combinations of Btu and sulfur content, we have differentiability and we are able to estimate stochastically our model, where the DTDF includes a measure of overall firm-specific TI and an idiosyncratic error term. We also directly control for endogeneity, interpret the partial effects of our MB constraints, compute firm-specific TI for each MB constraint, and calculate posterior densities for technical inefficiencies, productivity growth, and optimal directions. The DEA approach is non-stochastic and would not allow estimation of standard errors (unless bootstrapping is employed), overall TI as a stochastic residual jointly with idiosyncratic noise, estimation of TI for each MB equation, or the calculation of posterior densities. Further, we would not be able to deal with the endogeneity of inputs since DEA methods are non-stochastic.

To satisfy the MB constraints in (27) and (31), we incorporate two sets of equations into our DTDF system. We satisfy (27) by requiring that bad outputs are a linear function of good inputs, where for simplicity we temporarily include all pre-and post-combustion control devices in a $(1 \times D)$ vector $\tilde{\mathbf{x}}_a$ which we include along with $\mathbf{x}$ in a vector called $\mathbf{x}_c$. We also include bad inputs plus firm dummies (d_f) and plant dummies (d_p). Defining $R = N + D$ we write:

$$\begin{aligned} \tilde{y}_1 &= \textstyle\sum_{r=1}^{R} a_{r1}x_r + \sum_{j=1}^{J} a_{R+j,1}\tilde{x}_j + d_f + d_p + \varepsilon_1 + u_1 = \mathbf{a}_1'[\mathbf{x_c}', \tilde{\mathbf{x}}']' + d_f + d_p + \varepsilon_1 + u_1, \\ &\vdots \\ \tilde{y}_L &= \textstyle\sum_{r=1}^{R} a_{rL}x_r + \sum_{j=1}^{J} a_{R+j,L}\tilde{x}_j + d_f + d_p + \varepsilon_L + u_L = \mathbf{a}_L'[\mathbf{x_c}', \tilde{\mathbf{x}}']' + d_f + d_p + \varepsilon_L + u_L. \end{aligned} \tag{34}$$

We allow for one-sided error terms $u_1, ..., u_L \geq 0$. Idiosyncratic error terms $\varepsilon_1, ..., \varepsilon_L$ are included to capture measurement errors in the data. We interpret the u_l as the inefficiency with which each plant achieves allowable emissions given their set of inputs for emission control. Since u_l is defined as positive for the three pollutants in (46), for purposes of interpretation, a higher estimated u_l implies a higher level of uncontrolled emissions for the emissions MB constraints, using a given level of inputs. However, since a plant cannot overproduce bad outputs without having to pay huge fines, we assume that plants do not overproduce emissions. Thus, a larger u_l implies a larger TI for equation $l, l = 1, \ldots, L$ and means that additional good inputs must be expended to reduce emissions to the allowed levels.[15]

As elements of $\mathbf{x}_a$ in the NO_x MB equation, we include separate variables for the MW capacity associated with boilers that utilize the LNOX and OA pre-combustion technologies as well as a separate variable measuring the MW capacity whose post-combustion emissions are reduced using the SCR technology. For the SO_2 MB equation we replace the NO_x control-technology variables with ones measuring the SO_2 control technology: plant/firm dummies and a variable measuring the MW capacity associated with boilers whose emissions are scrubbed.

The first MB system in (34) can be written compactly as:

$$\tilde{\mathbf{y}} = \mathbf{A}[\mathbf{x_c}', \tilde{\mathbf{x}}']' + \mathbf{d}_f + \mathbf{d}_p + \boldsymbol{\varepsilon}_* + \mathbf{u}_*. \tag{35}$$

The second MB constraint in (31) can be compactly written as a system, given more that one bad input, as:

$$\tilde{\mathbf{x}} = \tilde{\mathbf{A}}\mathbf{x} + \mathbf{d}_f + \mathbf{d}_p + \boldsymbol{\varepsilon}_{**} + \mathbf{u}_{**}, \tag{36}$$

where $\mathbf{d}_f$ and $\mathbf{d}_p$ are vectors created by multiplying a $(L \times 1)$ unit column vector times d_t and d_t for (35) and a $(J \times 1)$ unit column vector times d_t and d_t for (36). Both equations are subject to the same type of two-component error. To interpret the one-sided component, consider (36) where a higher u_{**} for the sulfur MB equation implies that for a given level of good inputs, the plant must consume coal with a higher sulfur content. This translates into greater TI for this MB equation.

For all inputs and outputs we are either missing actual prices at the plant level (and use firm-level data instead) or we have at least some degree of measurement error with our plant level data. We have the prices of capital and labor at the firm rather than the plant level, so some measurement error may exist when we apply these prices to the plant. We have data on the price of energy (Btu from coal) at the plant level. However, these are prices for deliveries not consumption, which may cause slight discrepancies between reported and actual prices for consumption of coal within a calendar year. In rare cases good input and good output prices are missing and we generate them by interpolation. We have no data on prices of bad inputs and bad outputs. Thus, the entire vector of prices at a given time is $\mathbf{P}^*_t = [\mathbf{P}^*_{t1}, \mathbf{P}^*_{t2}]'$, where $\mathbf{P}^*_{t1}$ is the vector of latent prices for labor, capital, energy, and the good output, while $\mathbf{P}^*_{t2}$ is a vector of latent prices for bad inputs and bad outputs. For the former we have noisy data, while for the latter we have no data. Given the observed prices $\mathbf{P}_{t1}$ we assume a measurement-error model:

$$\mathbf{P}^*_{t1} = \mathbf{P}_{t1} + \boldsymbol{\xi}_{t1}, \tag{37}$$

where $\boldsymbol{\xi}_{t1}$ is a vector error term. For bad inputs and outputs, we assume that missing prices are related

[15]Hampf (2019) takes a different approach by assuming that abatement is unknown and is a econd component of the error term along with an idiosyncratic component.

to the observed prices and a set of exogenous variables:

$$\mathbf{P}^*_{t2} = \Gamma_o \mathbf{P}^*_{t1} + \Lambda_o \mathbf{r}_t + \boldsymbol{\xi}_{t2}, \tag{38}$$

where Γ_o and Λ_o are matrices of unknown coefficients (including a constant term that we omit for conciseness), while $\mathbf{r}_t$ includes regional and firm dummies, time dummies, and a restructuring dummy.[16] We specify the error vector as

$$\boldsymbol{\xi}_t = [\boldsymbol{\xi}'_{t1}, \boldsymbol{\xi}'_{t2}]' \sim N(0, \Psi). \tag{39}$$

We can now combine equations (37) and (38) as

$$\begin{bmatrix} I & 0 \\ -\Gamma_o & I \end{bmatrix} \mathbf{P}^*_t = \begin{bmatrix} \mathbf{P}_{t1} \\ \Lambda_o \mathbf{r}_t \end{bmatrix} + \begin{bmatrix} \boldsymbol{\xi}_{t1} \\ \boldsymbol{\xi}_{t2} \end{bmatrix}. \tag{40}$$

Defining $\begin{bmatrix} I & 0 \\ -\Gamma_o & I \end{bmatrix} = \Theta$ we have

$$\mathbf{P}^*_t = \Theta^{-1} \begin{bmatrix} \mathbf{P}_{t1} \\ \Lambda_o \mathbf{r}_t \end{bmatrix} + \Theta^{-1} \begin{bmatrix} \boldsymbol{\xi}_{t1} \\ \boldsymbol{\xi}_{t2} \end{bmatrix}. \tag{41}$$

Further, letting $\Theta^{-1} \begin{bmatrix} \boldsymbol{\xi}_{t1} \\ \boldsymbol{\xi}_{t2} \end{bmatrix} = \mathbf{v}_t$ this is a multivariate regression, so a Gibbs sampling step can be used to draw the non-zero elements of Γ_o and Λ_o, if we further assume that

$$[\epsilon_t, \mathbf{v}'_t]' \sim N(0, \Phi). \tag{42}$$

For Φ and Ψ we assume standard "non-informative" Wishart priors.

Given parameters $\Theta, \Gamma_o, \Lambda_o$, drawing $\mathbf{P}^*_t$ is performed using a Gibbs step when we combine (41) and (20)-(23). Once this step is performed the other parameters are drawn using a Hamiltonian Markov Chair Monte Carlo (MCMC) step by Girolami and Calderhead (2011) (GC). We provide additional details on this in on-line Appendix A-O.

3.2 The New MB Model

The New MB model consists of the DTDF in (17), the price equations in (20)–(23) (for $\mathbf{z} = \{\mathbf{x}, \mathbf{y}, \tilde{\mathbf{x}}, \tilde{\mathbf{y}}\}$), and two MB conditions in (35) and (36) for $\tilde{\mathbf{y}}$ and $\tilde{\mathbf{x}}$, subject to the translation property restrictions. Using $\mathbf{P}^*_t$, we then jointly estimate all of the equations of this model. One advantage of

[16] All variables are defined in the data section.

this is that we can obtain the coefficients in $\mathbf{A}$ and $\tilde{\mathbf{A}}$ plus the various "environmental" inefficiencies measured by $\mathbf{u} = (u, \mathbf{u}_{*}, \mathbf{u}_{**})$. To control for the potential endogeneity of $\mathbf{z}$, we employ as instruments the prices of good inputs, a restructuring dummy, a dummy for public/private ownership, and firm- and time-specific dummy variables.

3.3 Conventional Models

We estimate two forms of the Conventional model. Both eliminate the MB constraints and price equations (20)–(23) employed by the New MB model. The first Conventional model includes a standard DTDF, subject to the translation property restrictions, and reduced-form equations to model endogeneity:

$$\mathbf{z}_{(-1)} = \mathbf{\Pi f} + \boldsymbol{\varepsilon}_o, \tag{43}$$

where $\mathbf{z}_{(-1)}$ denotes a $(W-1)$ vector comprised of $\mathbf{z}$ excluding one element, say the first one (since this element is the dependent variable in (24)) and $\mathbf{f}$ is a $(d \times 1)$ vector of predetermined variables comprised of the same instruments as used by the New MB model. Further, $\mathbf{\Pi}$ is a $(W-1) \times d$ matrix. The second Conventional model is the same as the first Conventional model, except that it additionally excludes (43), which controls for endogeneity. This model is termed the Conventional model without Endogeneity.

3.4 New MB Models with Reduced Forms

We also estimate two versions of the New MB model with reduced forms. Both eliminate the price equations (20)–(23) employed by the New MB model. Instead, both versions include a standard DTDF, the MB constraints, and reduced-form equations to model endogeneity, subject to the translation property restrictions. Further, both allow for joint modeling of technical and environmental inefficiencies. The first version, termed the New MB model with Reduced Form, includes the following reduced-form equations:

$$\mathbf{z}_{(-1)} = \mathbf{\Pi f} + \boldsymbol{\varepsilon}_o, \tag{44}$$

where the predetermined variables are the same as in (43). This is like a standard Limited Information Maximum Likelihood (LIML) formulation *a la* Pagan (1979). For the idiosyncratic error terms we assume:

$$\boldsymbol{\varepsilon} = (\varepsilon, \boldsymbol{\varepsilon}_o, \boldsymbol{\varepsilon}_{*}, \boldsymbol{\varepsilon}_{**}) \sim \mathcal{N}_{N+2J+M+2L}\left(\mathbf{0}, \mathbf{\Sigma}\right). \tag{45}$$

The second version of this model, termed the New MB model with Lagged Values in the Reduced Form, adds lagged values of the explanatory variables in the reduced form.

Technical and environmental inefficiencies are modeled jointly as follows:

$$\mathbf{u} = (u, \mathbf{u}_{*}, \mathbf{u}_{**}) \sim \mathcal{N}^{+}_{L+J}\left(\mathbf{0}, \mathbf{\Omega}\right), \tag{46}$$

where $\mathcal{N}^{+}_{m}$ denotes the $m-$variate half-normal distribution.

4 Translation Property, Bayesian Priors, and Bayesian Analysis

4.1 Translation Property

To impose the translation property restrictions in (19), we first express them in the form:

$$\mathbf{\Delta}\left(\boldsymbol{\gamma}\right)\mathbf{g} = \boldsymbol{\zeta}, \tag{47}$$

where $\boldsymbol{\gamma} = (\gamma_w, \gamma_{ww'})$, $\mathbf{\Delta}\left(\boldsymbol{\gamma}\right)$ is a matrix containing the appropriate elements in (19) and $\boldsymbol{\zeta} \sim \mathcal{N}\left(\mathbf{m}, \omega_{\zeta}^2\mathbf{I})\right)$, where $\mathbf{m} = \begin{bmatrix} -1 \vdots \mathbf{0} \end{bmatrix}$ and we set ω_{ζ} to a small number. We have chosen $\omega_{\zeta} = 10^{-7}$. As (47) can be thought of as a *semi-informative prior* on $\boldsymbol{\gamma}$ conditional on $\mathbf{g}$, the prior is completed assuming a prior for $\mathbf{g}$:

$$\mathbf{g} \sim \mathcal{N}(\mathbf{0}, \omega_g^2\mathbf{I}). \tag{48}$$

A reasonable value for ω_g would be 1 subject to the restriction that $\mathbf{g}'\mathbf{g} = 1$. This imposes a normal prior for $\mathbf{g}$ on the unit hyper-sphere. Now write (47) in the form

$$\mathbf{\Psi}\left(\mathbf{g}\right)\boldsymbol{\gamma} = \boldsymbol{\zeta}, \tag{49}$$

and the DTDF in the form:

$$0 = \mathbf{Z}\boldsymbol{\gamma} + \varepsilon. \tag{50}$$

Here $\boldsymbol{\zeta} \sim N(0, \omega_g^2 I)$ and $\varepsilon \sim N(0, \sigma_o^2 I)$. In turn we have the system

$$\mathbf{m} = \mathbf{\Psi}\boldsymbol{\gamma} + \boldsymbol{\zeta}$$

$$0 = \mathbf{Z}\boldsymbol{\gamma} + \varepsilon$$

to which we apply GLS following Theil and Goldberger (1961) to obtain

$$\hat{\gamma}_{GLS} = \left(\left[\mathbf{\Psi}(\mathbf{g})'\ \mathbf{Z}'\right]\begin{bmatrix} 1/\omega_g^2 I & \\ & 1/\sigma_o^2 I \end{bmatrix}\begin{bmatrix} \mathbf{\Psi}(\mathbf{g}) \\ \mathbf{Z} \end{bmatrix}\right)^{-1}\left[(\mathbf{\Psi}(\mathbf{g})'\ \mathbf{Z}'\right]\begin{bmatrix} 1/\omega_g^2 I & \\ & 1/\sigma_o^2 I \end{bmatrix}\begin{bmatrix} \mathbf{m} \\ 0 \end{bmatrix}.$$

The GLS estimator of $\boldsymbol{\gamma}$ can be rewritten in the form

$$\hat{\gamma}_{GLS} = \left[\sigma_o^2 \boldsymbol{\Psi}(\mathbf{g})' \boldsymbol{\Psi}(\mathbf{g}) + \omega_g^2 \mathbf{Z}'\mathbf{Z}\right]^{-1} \boldsymbol{\Psi}(\mathbf{g})' \mathbf{m}. \tag{51}$$

4.2 Other Priors

Given our prior on $\boldsymbol{\gamma}$ and $\mathbf{g}$, the other priors are specified as follows:

$$p(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(dim(\boldsymbol{\Sigma})+1)/2}, \tag{52}$$

$$p(\boldsymbol{\Omega}) \propto |\boldsymbol{\Omega}|^{-(dim(\boldsymbol{\Omega})+1)/2}, \tag{53}$$

which are standard "non-informative" priors proposed by Zellner. For the coefficients of the reduced form:

$$\boldsymbol{\pi} = \text{vec}(\boldsymbol{\Pi}) \sim \mathcal{N}(\overline{\boldsymbol{\pi}}, \boldsymbol{\Sigma}_\pi). \tag{54}$$

We set $\overline{\boldsymbol{\pi}} = \mathbf{0}$ and $\boldsymbol{\Sigma}_\pi = \omega^2_{\boldsymbol{\Sigma}_\pi} \mathbf{I}$. In our baseline case we set $\omega_{\boldsymbol{\Sigma}_\pi} = 10^5$.

For the coefficients of the MB equations in (35) and (36) collectively denoted by $\mathbf{a}$ we assume:

$$\mathbf{a} \sim \mathcal{N}^+(\overline{\mathbf{a}}, \boldsymbol{\Sigma}_a), \tag{55}$$

where, in our baseline specification, $\overline{\mathbf{a}} = \mathbf{0}$ and $\Sigma_a = \omega^2_{\boldsymbol{\Sigma}_a} \mathbf{I}$ where $\omega_{\boldsymbol{\Sigma}_a} = 10^5 \mathbf{I}$. Notice that we have a multivariate half-normal, so we need to restrict these coefficients to be non-negative.

4.3 Bayesian analysis

As the model allows firm-specific directions, we have a hierarchical model whose posterior is quite cumbersome to explore efficiently. In this work we opt for a Girolami and Calderhead (2011) Hamiltonian Monte Carlo approach which explores first-order and second-order derivatives of the log posterior.

Moreover, we compute the marginal likelihood of the New MB model and the Conventional model using a Laplace approximation and the well-known candidate's formula (DiCiccio et al., 1997 and Lewis and Raftery, 1997). To ensure that results are robust we vary the baseline priors as follows:

i) We generate 10,000 random prior parameters.

ii) We re-estimate the models using sampling-importance-resampling.

In stage (i) the prior parameters are varied as follows:

$$\begin{aligned}
\overline{\mathbf{a}} &\sim \mathcal{N}(\mathbf{0}, 10^4\mathbf{I}), \\
\overline{\boldsymbol{\pi}} &\sim \mathcal{N}(\mathbf{0}, 10^4\mathbf{I}), \\
\frac{Q}{\omega_g^2} &\sim \chi^2(k),\ \ln Q \sim N(0, 10^2),\ k \sim Unif\left[10^{-7}, 30\right], \\
\frac{Q}{\omega_{\boldsymbol{\Sigma}\pi}^2} &\sim \chi^2(k),\ \ln Q \sim N(0, 10^2),\ k \sim Unif\left[10^{-7}, 30\right], \\
\frac{Q}{\omega_{\boldsymbol{\Sigma}a}^2} &\sim \chi^2(k),\ \ln Q \sim N(0, 10^2),\ k \sim Unif\left[10^{-7}, 30\right].
\end{aligned} \tag{56}$$

The last three priors are standard inverted *gamma* priors for the scale parameters involved in $\mathbf{g}, \boldsymbol{\Sigma}_\pi$ and $\boldsymbol{\Sigma}_a$. The scale parameters are supposed to arise from a fictitious sample of size k where Q is the sum of squares of "residuals". We are also interested in an overall measure of robustness for our key functions of interest that include directions, technical inefficiency, allocative inefficiency, returns to scale, and productivity growth. Suppose any of these five functions of interest is denoted by $f_{it,(p)}^{(s)}$ for observation i, t, MCMC draw $s = 1, ..., S$ and prior specification $p = 1, ..., P$, where $P = 10,000$. Suppose the baseline results are denoted by $f_{it,(0)}^{(s)}$. Our measure of robustness for the function of interest f is:

$$R_f = 100 \cdot (nTP)^{-1} \sum_{p=1}^{P} \sum_{i=1}^{n} \sum_{t=1}^{T} \left| \frac{f_{it,(p)}^{(s)} - f_{it,(0)}^{(s)}}{f_{it,(0)}^{(s)}} \right|. \tag{57}$$

5 Data

Our sample consists of the 80 largest coal-fired power plants in the US from 1995-2005. As listed in the on-line Appendix B-O, the majority of our sample plants are located in the Southern, Mid-Atlantic, or Midwestern states, with a few in the Rocky Mountain and Far Western states.[17] Ninety-one percent of them are in jurisdictions where rates are controlled by regulatory commissions. Hence, the price of electricity is determined by these commissions and demand is exogenously determined by consumers. In this case, a cost-minimization model would be appropriate. Alternatively, the plants' owners could exercise enough influence over the commission that they help determine price, so that quantity is endogenous. If so, a profit-maximization model might be appropriate. The other 9 percent of plants are not degegulated but are instead part of Regional Transmission Organizations (RTOs) or Independent Transmission Operators. These entities facilitate efficient transmission of power from producers to end users based on bid-pricing. Each utility provides the RTO with the MW capacity for each unit at a given bid-price per hour, which is based on operating costs.[18] These MW bid-price combinations form a supply curve for the RTO. The intersection of this curve with exogenously determined demand yields a single price of electricity. In the absence of collusion among suppliers, plants will produce so long

[17] A list of these plants and parent utilities are available from the authors upon request.

[18] Fixed costs are recovered either through rents earned in this auction or through a capacity-market bidding mechanism.

as their marginal cost is less than the exogenously determined price. Failure to do so would mean that a plant foregoes rents. Again, a cost-minimization model would be appropriate. However, some collusion and manipulation of electricity quantities by plants is a possibility. If so, a profit-maximization model might be more appropriate. For this paper, because of space constraints, we illustrate only the latter model. The cost-minimization model can easily be obtained by dropping the good output price equation.

For our DTDF, the technology modeled in this study consists of one good output, three bad outputs, three good inputs, and one bad input. The good output is net electrical generation measured in MWh, while the three bad outputs are the emissions in short tons of SO_2, NO_x, and CO_2. The three good inputs are the capital stock (measured in 1973 year dollars), the number of employees, and the heat content in millions of Btu of coal consumed at each power plant. The single bad input is the percent sulfur content (S) of the coal burned. While the power plants in our sample consume coal and either oil or natural gas, on average 99 percent of the Btu generated by each plant comes from coal consumption. We obtained our data from a number of different sources. The Federal Energy Regulatory Commission Form 1 provides labor and capital data for private electric power plants, while the EIA-412 survey is the source of this data for public power plants. While DOE halted the Energy Information Agency (EIA)-412 survey after 2003, the Tennessee Valley Authority voluntarily posted 2004-06 data for its electric power plants on-line. The U.S. Department of Energy (DOE) Form EIA-767 is the source of information about fuel consumption and net MWh generation by plant. The SO_2, NO_x, and CO_2 emissions data at the plant level come from the EPA as part of its Continuous Emissions Monitoring System. We obtain the sulfur content of the coal burned by plant from EIA Form 423.[19] Since an increasing number of private utilities did not report capital and labor data after 2005, we terminate our sample at this year.

As indicated in equation (32), we compute plant-level latent prices for capital, labor, and coal utilized in each year using firm-level prices for capital and labor in addition to the delivered price of coal to the plant. We compute the user cost of capital at the firm level using the corporate tax rate, the corporate property tax rate, the depreciation rate, the firm's yield, and the Handy-Whitman Index as in APT. The yield on the firm's latest issue of long-term debt comes from Moody's Public Utility Manual (before 2001) and from Mergent's Public Utility Manual after that time. From Form 1 we collect the wage paid by the firm as salaries plus wages for electric operating and maintenance workers divided by the number of workers for the firm. From EIA Form 423 we compute the price of coal per million Btu to the plant, using the price of the weighted-average monthly deliveries of coal. Although these deliveries are provided at the plant level, frequently 90-day stockpiles of coal are maintained. Thus, we do not have exact data on the price of the actual quantities of coal burned during a calendar year.

[19]We wish to thank Carl Pasurka for supplying us with the data on input and output quantities.

Using EIA Form 767, we generate a variable for the MW capacity associated with boilers using each of the two most popular pre-combustion NO_x controls. For each boiler we record the first, second, and third NO_x pre-combustion emission control strategy if it is LNOX or OA, which are by far the most widely adopted. We then map each boiler into its corresponding generator and record the amount of MW capacity that is associated with each pre-combustion technology. Then we aggregate this MW capacity to the plant level. We do the same for the amount of MW capacity associated with the most widely-adopted post-combustion technology, the SCR control method.[20] Finally, we record the cumulative MW capacity whose associated emissions each plant scrubs to remove SO_2. We standardize all data except for dummy variables to eliminate units of measurement in our calculation of optimal directions of movement to the DTDF.

6 Empirical Results

In Table 1 we use the baseline prior to compute the log Bayes Factors (BF) in favor of the New MB model against the Conventional model. The BF in favor of the New MB model is 1285.44 and the 95% BF ranges from 977.16-1588.30 across different models. Based on 10,000 alternative priors, we also provide robustness checks in terms of percent changes for $\boldsymbol{g}$, TI, allocative inefficiency (AI), returns to scale (RTS), and productivity growth (PG) over these alternative priors. The table results indicate that these measures are not sensitive to the alternative priors. We define RTS as

$$\text{RTS} = -\frac{\nabla_{\text{x}} \overrightarrow{D}(\mathbf{y}, \mathbf{x}^{*})\mathbf{x}}{\nabla_{\text{y}} \overrightarrow{D}(\mathbf{y}, \mathbf{x}^{*})\mathbf{y}}, \tag{58}$$

where RTS > 1, RTS $= 1$, and RTS < 1 imply increasing, constant, and decreasing returns to scale, respectively.

We compute PG and TI as in Atkinson and Tsionas (2016). We examine two measures of TI. The first is overall TI, which measures the extent to which a firm fails to produce the frontier output for a given set of inputs, where TI is normalized to lie in the unit interval, so that the frontier firm has a TI=0. The second type of TI is computed in an analogous manner for each MB constraint. For overall TI, the change in TI over time is defined as efficiency change (EC). The shift in the production frontier over time is defined as technical change (TC). Then PG= EC + TC.

The AI of a firm increases as it fails to equate each input price to its marginal value product. We calculate AI for the DTDF only regarding inputs and outputs for which we have prices. Suppose these inputs and outputs are denoted by $\mathbf{X}$, their fitted values by $\hat{\mathbf{X}}$, and their prices by $\mathbf{P}^{*}$. Our measure of

[20]Only one boiler utilized a selective non-catalytic reduction technology.

AI is computed as

$$\mathrm{AI}_{it} = \mathcal{D}_{it} - \min_{i,t} \mathcal{D}_{it}, \tag{59}$$

where

$$\mathcal{D}_{it} = \frac{\mathbf{P}^{*\prime}_{it}(\mathbf{X}_{it} - \hat{\mathbf{X}}_{it})}{\mathbf{P}'_{it}\mathbf{X}_{it}}. \tag{60}$$

Table 1: Log Bayes factors in favor of New model, against the Conventional model, and robustness measures

Baseline prior, BF	1285.44
95% BF Range	977.16 – 1588.3
Robustness, $\mathbf{g}$, (%)	0.006
Robustness, TI (%)	0.003
Robustness, AI (%)	0.002
Robustness, RTS (%)	0.002
Robustness, PG (%)	0.002

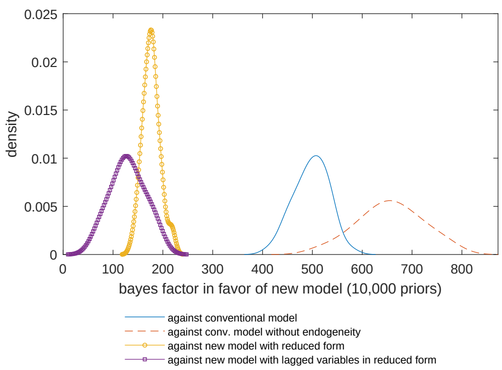
In equations (34) we have assumed over-production of bad outputs by adding one-sided error terms to the quantity equations for bad outputs. In Table 2 we report the Bayes factors in favor of this version of the New MB model relative to other specifications of this model which assume various types of under-production of bad outputs. We find that the assumption of over-production is clearly superior.

Table 2: Bayes factors in favor of present model

$-w \geq 0$ in all equations	417.12
$u - w, u, w \geq 0$ in all equations	282.16
$-w \geq 0$ in CO2 eq. only	203.75
$-w \geq 0$ in SO2 eq. only	324.50
$-w \geq 0$ in NOx eq. only	355.44
$u - w \geq 0$ in CO2 eq. only	437.70
$u - w \geq 0$ in SO2 eq. only	613.52
$u - w \geq 0$ in NOx eq. only	562.14

In Figure 2 we report BF from 10,000 alternative priors of the New MB model (with the latent price equations and the MB constraints) against certain other models. The first is the Conventional model (where endogeneity is modeled with a reduced form instead of the first-order conditions and MB constraints are excluded). The second is the Conventional model without Endogeneity. The third is the New MB model with the Reduced Form (44) instead of the first-order conditions in (20)-(23). The last is the New MB model when lagged values of the explanatory variables are also included in the reduced form (44). The smallest BF for any comparison of models is 20, which occurs for the New MB model compared to the Conventional model. All other comparisons yield posterior means that are as much as four times larger. This occurs for the New MB model compared to the Conventional model without Endogeneity. The results from Table 2 and Fig. 2 strongly support the New MB model to deal with endogeneity against a variety of alternatives for the New MB and Conventional models. In on-line Appendix C-O, we provide additional out-of-sample analysis that is consistent with the results in Figure 2.

Figure 2: Bayes factors in favor of the New MB model against all other models



Notes: New MB model includes first-order conditions in (20)-(23) with MB constraints. Conventional model includes the reduced form in (44). Conventional model without endogeneity excludes the reduced form equation (44). Both eliminate MB constraints and price equations. New MB model with reduced form uses (44) instead of the first-order conditions in (20)-(23) with MB constraints. New MB model with lagged variables in reduced form, includes lagged values of all input and outputs in the reduced form (44) with MB constraints.

In Table 3 we present posterior means and posterior standard deviations of the estimated MB equations for the New MB model. Columns 2-4 relate each bad output to the good inputs, the bad input (sulfur), the control technologies, and dummy variables. Column 5 relates sulfur to the good inputs and dummy variables. Consistent with (**??**) for all equations the coefficients for pre- and post-combustion control of SO_2 and NO_x are all negative with posterior means located far from zero. Consistent with (29) the coefficients of all good explanatory variables for S are positive with posterior means located far from zero. These coefficients provide the weights for a unit of each explanatory variable to increase or decrease CO_2, SO_2, NO_x, and S. Since these variables are measures of MW capacity covered by each technology, we can compare the magnitude of estimated coefficients. The use of the low-NO_x-overfire-air technology reduces emissions about 7% more than the low-NO_x-combustion technology. The use of the post-combustion SCR technology is only slightly less efficient than this technology.[21] The marginal effects of each pre- and post-combustion control for NO_x are roughly equal, indicating that the plant has roughly equalized the productivity of each of these devices.

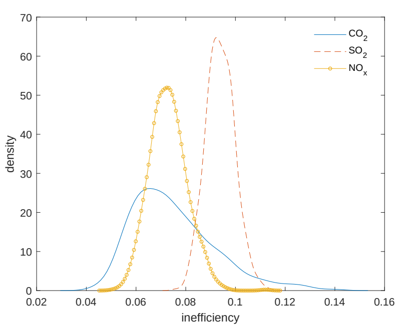
[21] Since boilers employing Low-NO_x are linked to a capacity of 56,955 MW/year on average and those employing Overfire Air are linked to capacity of 23,482 MW/year on average, the former method is far more important in controlling emissions.

Table 3: Material Balance Equations

Explanatory Variables	CO_2	SO_2	NO_x	S
Capital	0.213	0.094	0.144	0.181
	(0.025)	(0.012)	(0.018)	(0.016)
Labor	0.017	0.044	0.057	0.062
	(0.005)	(0.012)	(0.014)	(0.012)
Heat content (coal)	0.506	0.515	0.482	0.544
	(0.012)	(0.016)	(0.017)	(0.022)
Sulfur	0.142	0.135	0.155	—
	(0.015)	(0.016)	(0.013)	
Low -NO_x combustion boiler	—	—	–0.188	—
			(0.022)	
Low-NO_x Overfire Air	—	—	–0.203	—
			(0.014)	
Selective Catalytic Reduction	—	—	–0.199	—
			(0.023)	
Flue Gas Desulfurization	—	–0.184	—	—
		(0.032)		
Firm Dummies	Yes	Yes	Yes	Yes
Plant Dummies	Yes	Yes	Yes	Yes

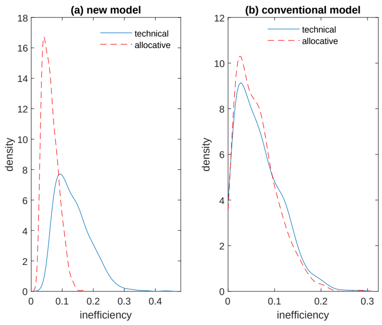
We estimate TI for each MB equation and present the results in Figure 3 for CO_2, SO_2, and NO_x for the New MB model. Although the mean posterior inefficiencies are roughly identical for all pollutants (about .04-.05), the density for SO_2 is slightly more concentrated than the densities for the other two pollutants. However, these inefficiencies could be reduced by improving plant management.

Figure 3: Technical Inefficiency for Material Balance Equations: New MB model



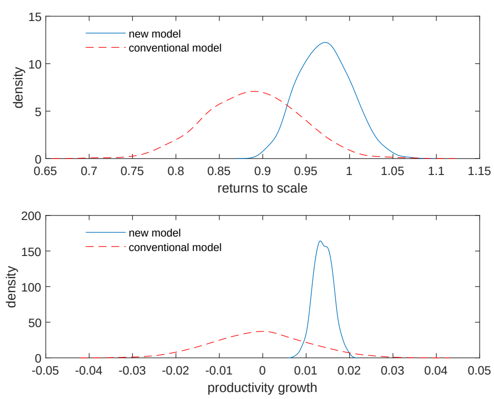
In Figure 4, we report TI and AI for the DTDF produced by the Conventional and New MB models. With both models, TI and AI are quite small. For the Conventional model, both measures have posterior means equal to about .03. For the New MB model, AI has roughly the same posterior mean, while that for TI is about .1.

Figure 4: Technical and Allocative Inefficiency:
New MB model versus Conventional model



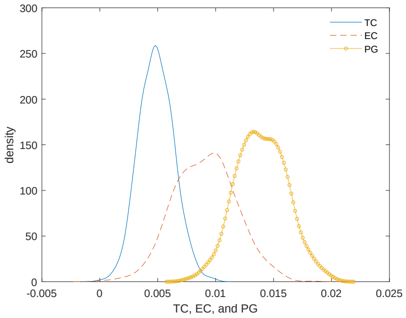
In Figure 5 posterior densities for RTS and PG differ substantially for the same two models. We calculate posterior means for RTS of about .96 for the New MB model, but only approximately .88 for the Conventional model. Thus, the New MB model indicates that the sample plants are operating closer to the range of constant returns to scale. In terms of PG the New MB model yields a posterior density that is much less disperse and has a higher mean, approximately equal to .015, than the Conventional model, whose posterior mean is about .005.

Figure 5: Returns to Scale and productivity growth: New MB model versus Conventional model



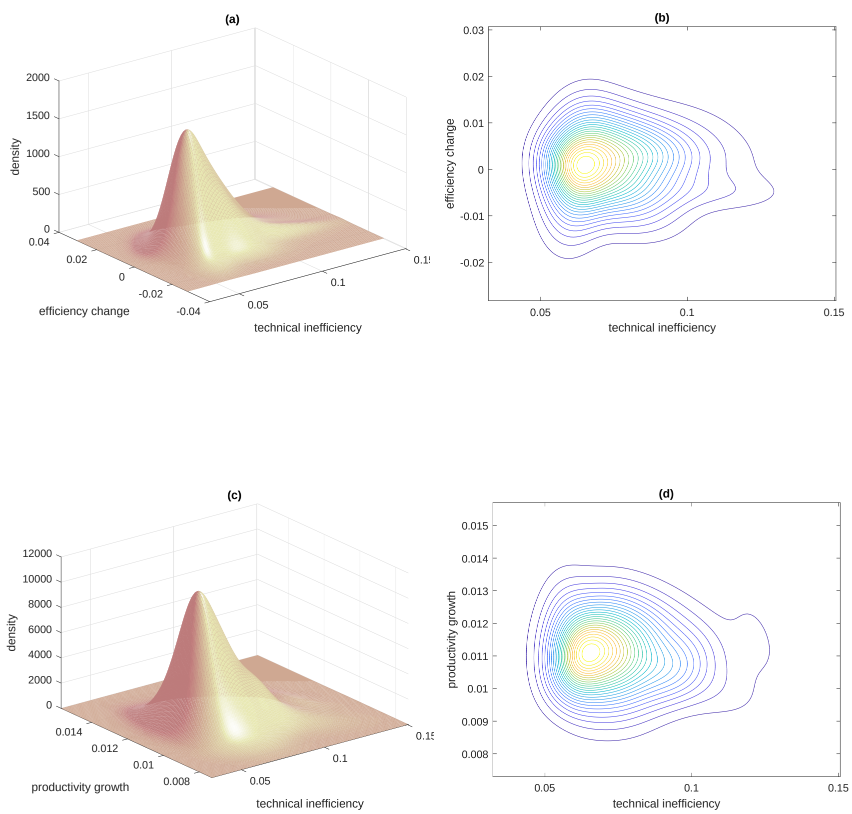
In Figure 6 posterior densities for EC, TC, and PG for the preferred New MB model are small but positive. The average posterior for TC is about .005, that for EC is about .01, while that for PG is about .015. That is, the outward movement of the production frontier over time is very small, but catching up to the frontier is about twice the outward movement, so that PG is about 1.5 %. However, this is small for a large capital-intensive industry such as coal-fired electricity generation and is most likely due to minor investments in more efficient generation capital over time.

Figure 6: EC, TC, PG for New MB model



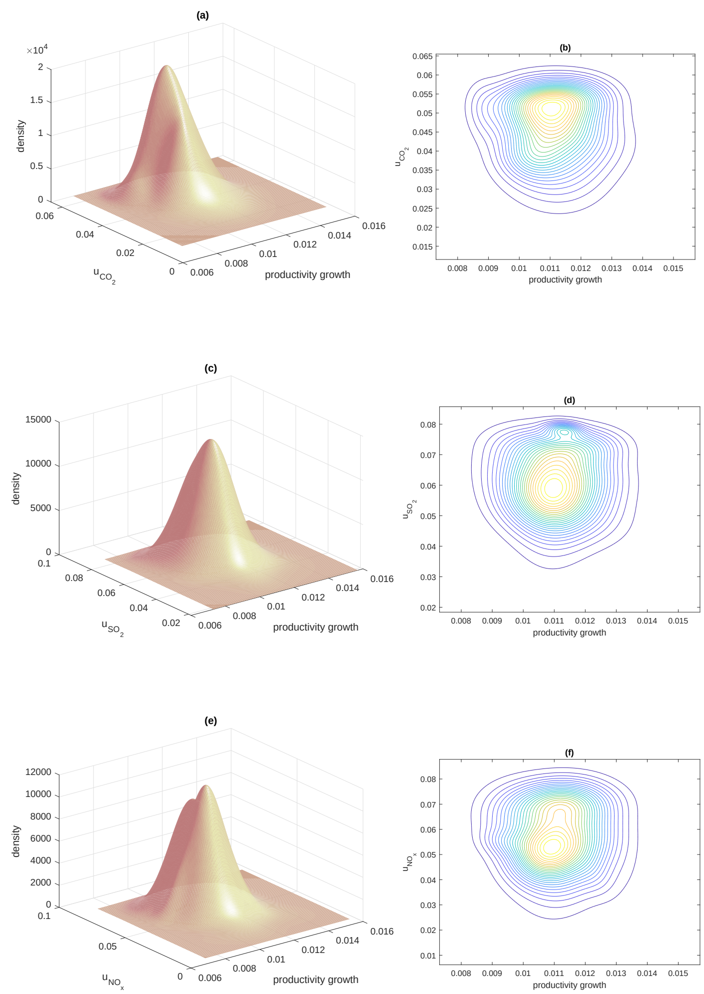
In Figure 7 we report bivariate densities (panels (a) and (c)) and contours (panels (b) and (d)) for EC versus TC (upper part) and PG versus TI (bottom part). Both relationships are uni-modal, but their elongated forms suggest that many plants may be able to lessen TI and improve EC and PG, or at least reduce TI without worsening EC or PG. This indicates the need for better management practices to address the problem of low efficiency.

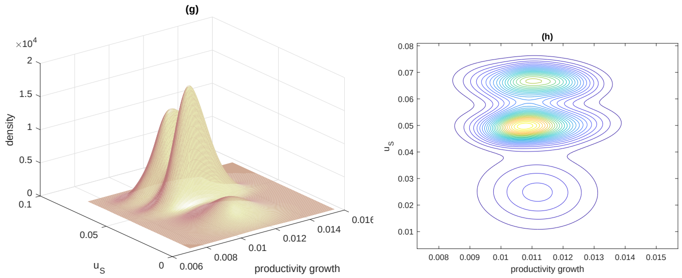
Figure 7: Bivariate posteriors of efficiency change, technical inefficiency, and productivity growth



In Figure 8 we report posterior bivariate densities for PG and TI for each of our MB equations. In panels (a) and (c) we see that u_{CO_2} and u_{SO_2} have uni-modal bi-variate distributions with respect to PG. However, some firms can decrease u_{CO_2} and u_{SO_2} while maintaining or increasing PG at the same time. In panels (e) and (g) we observe bi-modal posterior distributions for u_{NO_x} and u_S, each relative to PG. For panel (e), there is substantial room for plants to reduce TI for NO_x and increase PG at the same time. For panel (g), opportunities for improvement are even more dramatic. The two distinct clusters indicate that plants in the upper cluster can substantially decrease u_S for a given level of PG by shifting to a lower cluster, or they can increase PG for a given level of u_S within a cluster. Improved management skills or better technologies may be required to achieve these changes.

Figure 8: Aspects of bivariate posteriors of productivity growth and one-sided error terms in the MB equations





In Appendix A we present a number of additional bivariate posterior densities relating TI from individual MB equations to each other and to TI and AI. Results reinforce the previous evidence that improvemts in efficiency can be attained by many plants without sacrifice, through improved managerial methods.

We report posterior moments for overall TI and the u terms for each MB equation in Table 4. The posterior mean for overall TI, $u_{\mathrm{CO_2}}$, and $u_{\mathrm{NO_x}}$ are roughly .07, while that for $u_{\mathrm{SO_2}}$ is roughly twice this size. The posterior mean for u_S is .052. The simple pairwise correlations of $u_{\mathrm{CO_2}}, u_{\mathrm{SO_2}}, u_{\mathrm{NO_x}}$, and u_{S} are at least .90 across all firms and time periods.

Table 4: Posterior moments for Overall TI and all MB Inefficiency Terms

	post. mean	post. s.d
Overall TI	0.075	0.017
$u_{\mathrm{CO_2}}$	0.073	0.0074
$u_{\mathrm{SO_2}}$	0.135	0.055
$u_{\mathrm{NO_x}}$	0.062	0.024
u_{S}	.052	.015

In Figures 9-11 we report posterior means for optimal directions of good inputs and the good output for the New MB model and the Conventional model. We observe considerable differences between the two models in all figures. Since the new model is our preferred model, we only discuss these results in detail. The typical plant, in moving from its current position to its profit-maximizing position, should keep labor at current levels, slightly decrease capital, output, and S, more substantially reduce CO_2, yet moderately increase SO_2 and NO_x and more substantially increase energy. The need to substantially decrease CO_2 is due to our estimation of the plant's profit-maximizing position subject a positive latent price for such emissions, which did not exist during our sample period. If managers anticipate a cap-and-trade market for CO_2, they would be well-advised to make such changes. The increase in NO_x production is most likely in response to the adoption of pre-control devices that operate at fixed levels, thereby over-controlling this pollutant. The same is true for SO_2, where a myriad of state and local regulations required that plants adopt specific boiler technologies.

Figure 9: Optimal Directions–Good Inputs and Good Output: New MB model versus Conventional model

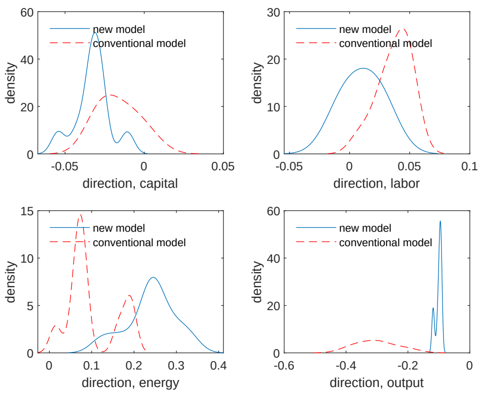


Figure 10: Optimal Directions-Bad Outputs: New MB model / versus Conventional model

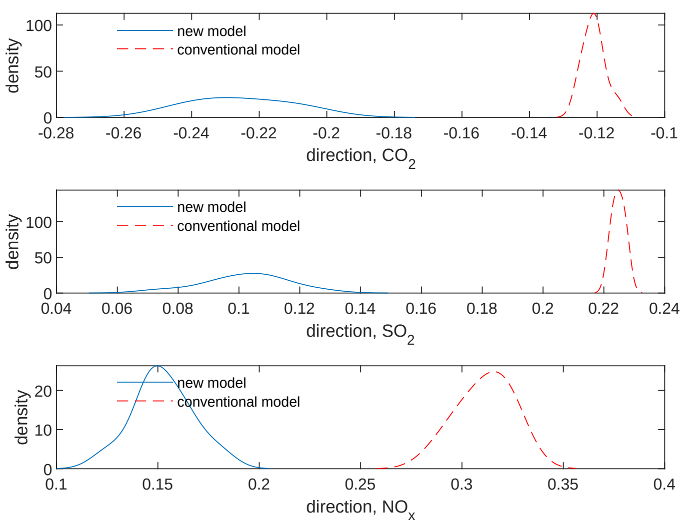
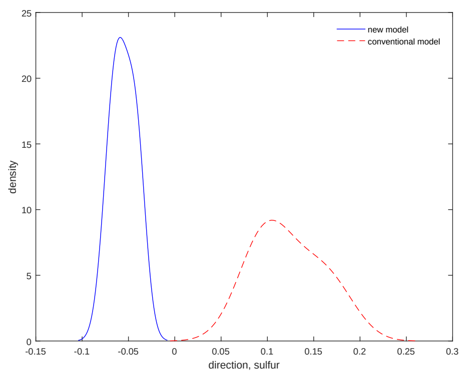


Figure 11: Optimal Directions-Bad Input: New MB model versus Conventional model



7 Conclusion

The DTDF has been employed in an SFA context with increasing frequency to more flexibly estimate multiple-input and multiple-output production, firm inefficiency, and productivity growth without employing MB constraints. Our paper is the first to introduce MB constraints into this estimation model, which we term the New MB model. Subject to these constraints, we compute optimal directions assuming plant profit maximization, correct for the potential endogeneity of all inputs and outputs, subject to latent prices for regulated bad outputs and the bad input. The most important innovation of this paper is our utilization of MB constraints when modeling the production of bad outputs and the use of bad inputs. We show how to incorporate stochastic MB constraints in the modeling of the production process so that tradeoffs between inputs and outputs (both good and bad) satisfy production and abatement constraints. We incorporate pre- and post-combustion technologies for control of NO_x and SO_2. In so doing, we satisfy the first and second laws of thermodynamics.

We illustrate these methods. using a balanced panel of 80 coal-fired U.S. electric utility plants from 1995-2005. Bayes factors indicate that the New MB model is clearly superior to the Conventional model which omits MB constraints. We observe substantial differences between the specifications that include the MB constraints and those that omit them in terms of estimated productivity growth and efficiencies. Specifically, compared to the Conventional model, the New MB model indicates less technical and allocative inefficiency, a greater returns to scale, higher productivity growth, and different optimal directions. In many cases, the posteriors for these measures become more precise with the New MB model. Analysis of bivariate posterior densities for the error terms of the MB equations indicates that plant managers appear capable of making many improvements in pollution control efficiency without sacrificing productivity growth or overall plant efficiency.

References

Archibald, S.O. (1988). Incorporating externalities into agricultural productivity analysis. In: Capalbo, S.M., Antle, J.M. (Eds.), Agricultural Productivity: Measurement and Explanation. Resources for the Future: Washington, 1988, pp. 366–393.

Atkinson, S. E., Primont, D., & Tsionas, M.G. (2018). Statistical inference in efficient production with bad inputs and outputs using latent prices and optimal directions. Journal of Econometrics, 204 (2), 131-146.

Atkinson, S. E. & Tsionas, M. G. (2016). Directional Distance Functions: Optimal Endogenous Directions, Journal of Econometrics, 190, 301-314.

Ayres, R.U. & Kneese, A.V. (1969). Production, consumption, and externalities. American Economic Review, 69, 282-296.

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

Chambers, R. G. (1998). Input and output indicators, in R. Färe, Grosskopf, S., & Russell, R., eds., Index Numbers: Essays in Honor of Sten Malmquist Boston: Kluwer Academic Publishers , 241-271.

Chambers, R.G., Chung, Y., & Färe, R. (1998). Profit, directional distance functions and Nerlovian efficiency. Journal of Optimization Theory and Applications, 98, 351-364.

Charnes, A. , Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. European Journal of Operational Research, 2 (6), 429–444.

Chung Y.H., Färe, R., & Grosskopf, S. (1997). Productivity and undesirable outputs: A directional distance function approach Journal of Environmental Management, 51 (3): 229-240

Coelli, T., Lauwers, L., & Van Huylenbroeck, G. (2007). Environmental efficiency measurement and the materials balace condition. Journal of Productivity Analysis, 28 (1-2), 3–12.

Cooper, W., Seiford, L. M., & Zhu, J. (Eds.) (2011). Handbook on data envelopment analysis (164). Springer Science & Business Media.

Dakpo, K.H., Jeanneaux, P., & Latruffe, L. (2015). Modelling pollution-generating technologies in performance benchmarking: Recent developments, limits and future prospects in the nonparametric framework. European Journal of Operational Research, 250 (2), 347-359.

De Koeijer, T.J., Wossink, G.A.A., van Ittersum, M.K., Struik, P.C., & Renkema, J.A. (1999). A conceptual model for analysing input–output coefficients in arable farming systems: from diagnosis towards design. Agricultural Systems, 61, 33–44.

De Koeijer, T.J., Wossink, G.A.A., Struik, P.C., & Renkema, J.A. (2002). Measuring agricultural sustainability in terms of efficiency: the case of Dutch sugar beet growers. Journal of Environmental Management, 66, 9–17.

Department of Energy, Energy Information Administration (EIA) Form 767 Boiler Files, 1988-1997, Washington, D.C.

DiCiccio, T.J., Kass, R.E., Raftery, A., & Wasserman, L. (1997). Computing Bayes factors by combining simulation and asymptotic approximations. Journal of the American Statistical Association, 92 , 903-915.

Ebert, U. & Welsch, H. (2007). Environmental Emissions and Production Economics: Implications of the Materials Balance American Journal of Agricultural Economics, Vol. 89, No. 2 (May, 2007), pp. 287-293

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

Färe, R. & Grosskopf, S. (2004). Modeling undesirable factors in efficiency evalua- tion: comment. European Journal of Operational Research, 157 (1), 242–245.

Färe, R., Grosskopf, S., & Pasurka, C. (2007a). Environmental production functions and environmental directional distance functions. Energy, 32 (7), 1055–1066.

Färe, R., Grosskopf, S., & Pasurka, C. (2007b). Pollution abatement activities and traditional productivity, Ecological Economics, 62 (3), 673–682.

Federal Energy Regulatory Commission Form 1: Annual Report of Major Electric Utilities, 1988-2005.

Førsund, F.R. (2018). Multi-equation modelling of desirable and undesirable outputs satisfying the materials balance. Empirical Economics, 54, 67-99.

Girolami, M. & Calderhead, B. (2011). Riemann manifold Langevin and Hamiltonian Monte Carlo methods. Journal of the Royal Statistical Society B, 73, 123-214.

Hailu, A. & Veeman, T.S. (2001). Alternative methods for environmentally adjusted productivity analysis. Agricultural Economics, 25, 211-218.

Hampf, B. (2014). Separating environmental efficiency into production and abatement efficiency: a nonparametric model with application to US power plants. Journal of Productivity Analysis, 41, 457-473.

Hampf, B. (2019). Estimating emission coefficients and mass balances using economic data: A stochastic frontier approach. Journal of industrial ecology, 23, 932- 945.

Hampf, B. & Rødseth, K. L. (2015). Carbon dioxide emission standards for US power plants: An efficiency analysis perspective. Energy Economics, 50 , 140–153.

Hill, S.B., Vincent, C., & Chouinard, G. (1999). Evolving ecosystems approaches to fruit insect pest management. Agriculture Ecosystems and Environment, 73, 107–110.

Hudgins, L. B. & Primont, D. (2007). Derivative properties of directional technology distance functions, in Färe, R. Grosskopf, S., and Primont, D. eds., in Aggregation, Efficiency, and Measurement, Springer.

Lauwers, L. (2009). Justifying the incorporation of the materials balance principle into frontier-based eco-efficiency models. Ecological Economics,68,16051614.

Lewis, S.M. & Raftery, A. E. (1997). Estimating Bayes factors via posterior simulation with the Laplace-Metropolis estimator. Journal of the American Statistical Association, 92, 648-655.

Moody's Public Utility Manual, Moody's Investors Service, 1988-2005, New York.

Murty, S. M., Russell, R. R., & Levkoff, S. B. (2012). On modeling pollution-generating technologies. Journal of Environmental Economics and Management, 64, 117-135.

Pagan, A. (1979). Some consequences of viewing LIML as an iterated Aitken estimator. Economics Letters, 3, 369-372.

Pethig, R. (2006). Non-linear production, abatement, pollution and materials balance reconsidered. Journal of Environmental Economics and Management, 51, 185– 204.

Picazo-Tadeo, A., & Prior, D. (2009). Environmental externalities and efficiency measurement. Journal of Environmental Management, 90, 3332-3339.

Prasad, S., & Calis, A. (1999). Capability indices for material balance accounting, European Journal of Operational Research 114 (1), 93–104.

Rennings, K. (2000). Redefining innovation-eco-innovation research and the contribution from ecological economics. Ecological Economics, 32, 319–332

Rødseth, K. L. (2016). Environmental efficiency measurement and the materials balance condition reconsidered. European Journal of Operational Research 250 (1), 342–346

Sahoo, B. K., Luptacik, M., & Mahlberg, B. (2011). Alternative measures of environmental technology structure in DEA: An application. European Journal of Operational Research, 215 (3), 750–762.

Scheel, H. (2001). Undesirable outputs in efficiency valuations. European Journal of Operational Research, 132 (2), 400–410.

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

Seiford, L. M. & Zhu, J. (2005). A response to comments on modeling undesirable factors in efficiency evaluation. European Journal of Operational Research, 161 (2), 579–581.

Sueyoshi, T., & Goto, M. (2012). Data envelopment analysis for environmental as- sessment: Comparison between public and private ownership in petroleum in- dustry. European Journal of Operational Research, 216 (3), 668–678.

Theil, H., & Goldberger, A. S. (1961). On Pure and Mixed Statistical Estimation in Economics. *International Economic Review* 2, No. 1, 65-78.

Van Meensel, J., Lauwers, L., Van Huylenbroeck, G., & Van Passel, S. (2010). Comparing frontier methods for economic-environmental trade-off analysis. European Journal of Operational Research, 207, 1027–1040.

Wang, K., Wei, Y.-M., & Zhang X. (2012). A comparative analysis of China's regional energy and emission performance: Which is the better way to deal with undesirable outputs?, Energy Policy, 46, 574–584.

Wang K., Wei,Y.-M, & Zhang X. (2013). Energy and emissions efficiency patterns of Chinese regions: A multi-directional efficiency analysis, Applied Energy, 104, 105–116.

Wang, K., Wei, Y.-M., & Huang, Z. (2018). Environmental efficiency and abatement efficiency measurements of Chinas thermal power industry: A data envelopment analysis based materials balance approach. European Journal of Operational Research, 269 (1), 3550.

Wang H., Zhou, P., Xie, B-C.,& Zhang, N. (2019). Assessing drivers of CO2 emissions in China's electricity sector: A metafrontier production-theoretical decomposition analysis. European Journal of Operational Research, 275 (3), 1096-1107.

Welch, E., & Barnum, D. (2009). Joint environmental and cost efficiency analysis of electricity generation. Ecological Economics, 68 (8), 2336–2343.

Wossink, G.A.A., Lansink, A.G.J.M.O., & Struik, P.C. (2001). Non-separability and heterogeneity in integrated agronomic-economic analysis of nonpoint-source pollution. Ecological Economics 38, 345–357

Zhu, J. (2014). Quantitative models for performance evaluation and benchmarking: Data envelopment analysis with spreadsheets. Springer International Publishing Switzerland.

8 Appendix: Additional Bivariate Posterior Densities

In Figures A1-A3, plants exhibit a strong pairwise correlation of the u for each MB equation. This means that managers that are inefficient (efficient) in the control of one pollutant may be inefficient (efficient) in the control of all pollutants. Reductions in TI for one pollutant will be achieved jointly with another pollutant. The existence of a smaller second cluster indicates that some firms may be able to make additional improvements by shifting to the lower cluster.

Figure A1: Aspects of bivariate posteriors of $u_{\mathrm{SO_2}}$ and $u_{\mathrm{CO_2}}$ in MB equations

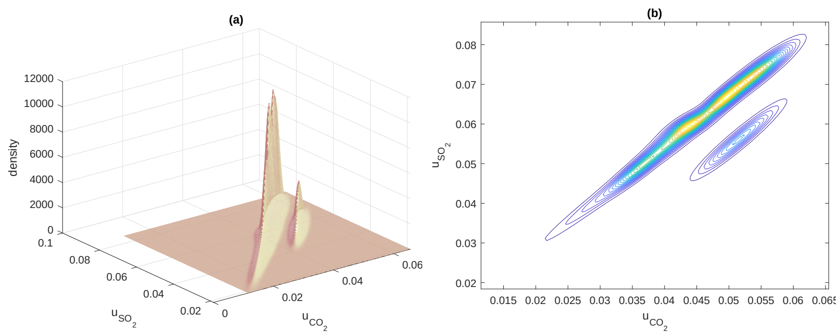


Figure A2: Aspects of bivariate posteriors of u_{NO_x} and $u_{\mathrm{CO_2}}$ in MB equations

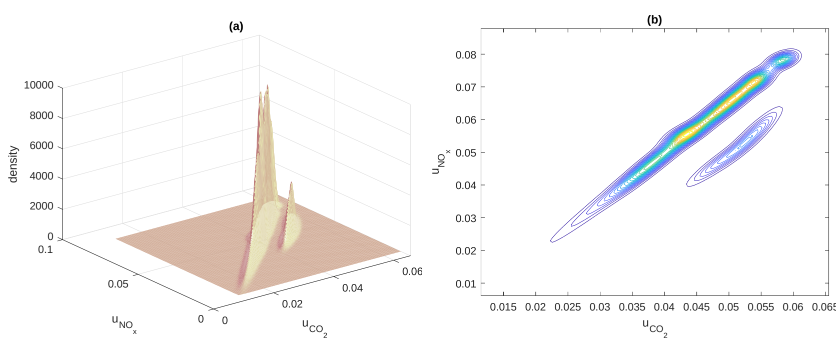
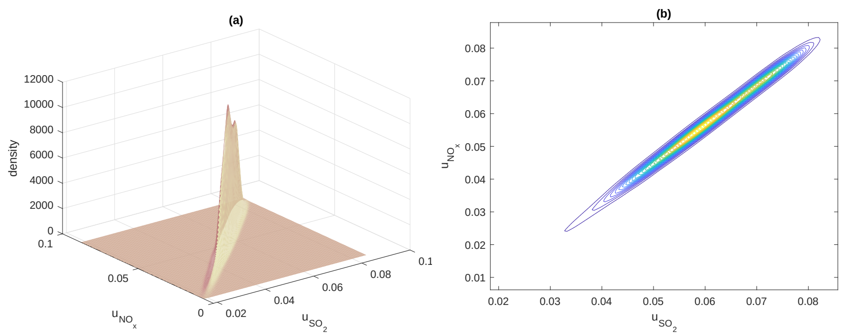


Figure A3: Aspects of bivariate posteriors of u_{NO_x} and u_{SO_2} in MB equations



Figures A4-A6 present bivariate posterior densities for u_{S} with u_{CO_2}, u_{SO_2}, and u_{NO_x}. Compared to Figures A1-A3, there are many more clusters in Figures A4-A6. Usually two modes on each of two clusters are visible in each of these figures. There is also a slightly negative relationship for each of the pairwise comparisons, locally in the vicinity of the different modes. This suggests the existence of a moderate tradeoff, so that reducing TI in the sulfur MB equation results in an increase in TI for the three emission MB equations. This is consistent with a disconnect between the optimal choice of sulfur in the production function and the TI of pollutant removal. Again, the stacked clusters indicate substantial potential reductions in the TI for sulfur for a given level of TI for each pollutant, when moving downward to a lower cluster in these figures. More attention needs to be paid to the interaction of production and emission control. Managers should work to more closely integrate the two tasks so that any negative tradeoff is eliminated.

Figure A4: Aspects of bivariate posteriors of u_{S} and u_{CO_2} in MB equations

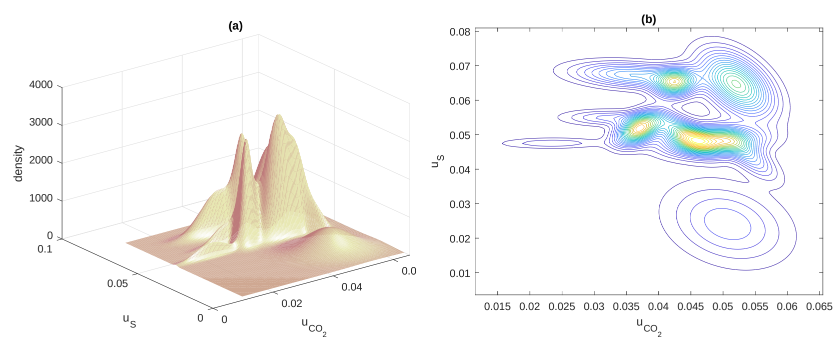


Figure A5: Aspects of bivariate posteriors of u_S and u_{SO_2} in MB equations

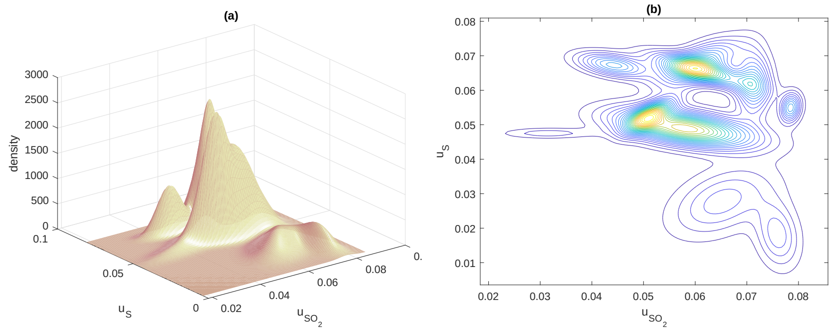
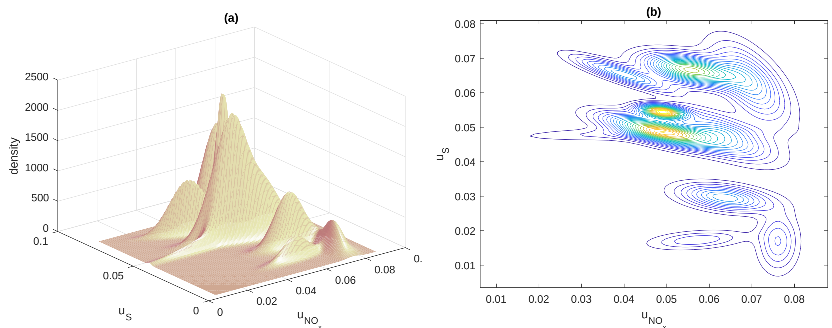


Figure A6: Aspects of bivariate posteriors of u_S and u_{NO_x} in MB equations



Another interesting managerial issue concerns the bivariate posterior relationship between overall TI and the TI terms in the MB equations, which we examine in Figure A7. All panels indicate substantial potential to reduce both inefficiencies in each bivariate plot. Panel (g) relates u_S to TI and indicates that plants can reduce u_S substantially for a given level of TI by moving to a lower cluster. Further, considerable opportunities exist to improve overall TI for a given level of u_S. Both improvements would require adopting better management practices or technologies.

Figure A7 Aspects of bivariate posteriors of technical inefficiency and one-sided error terms in the MB equations

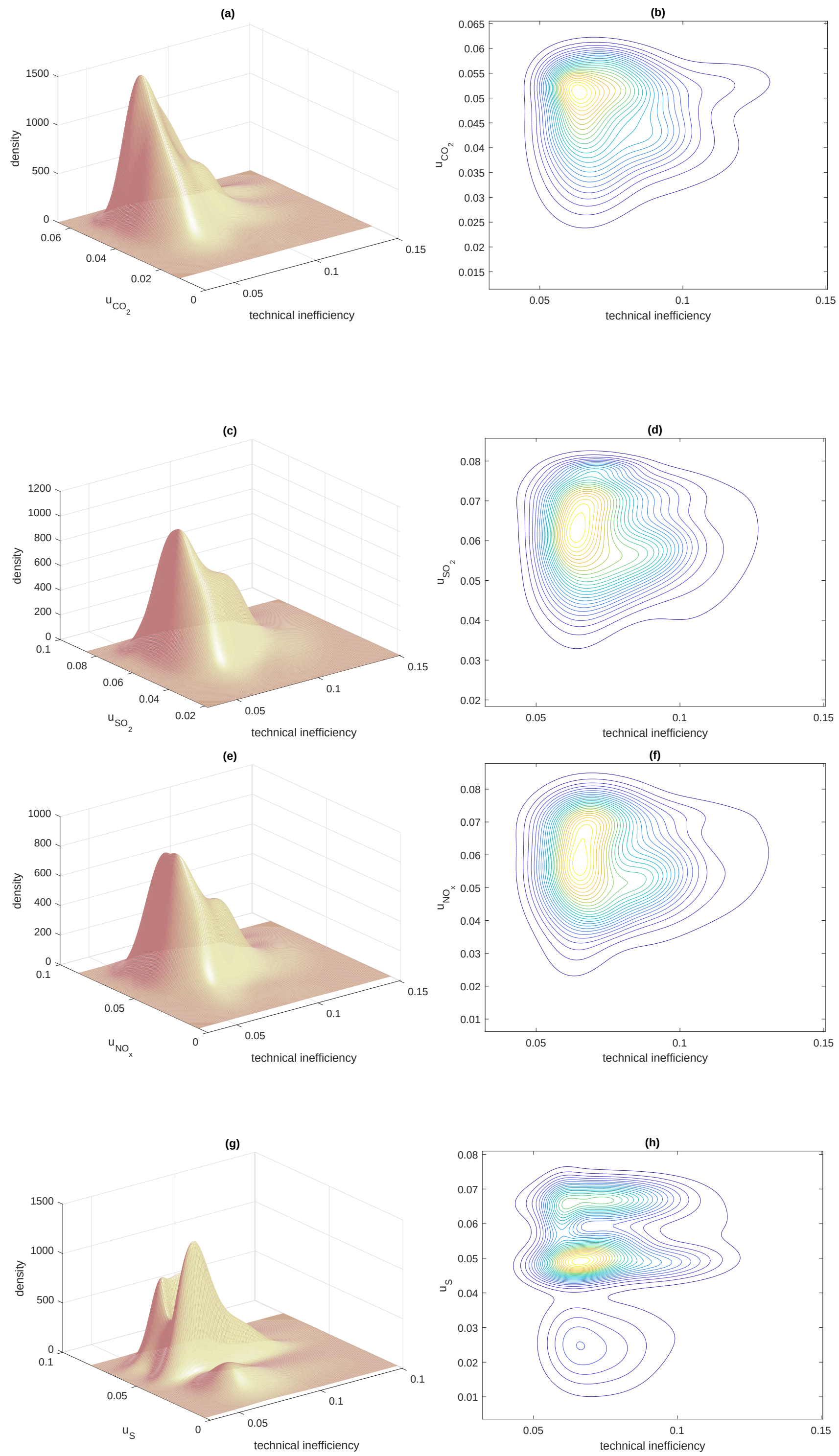
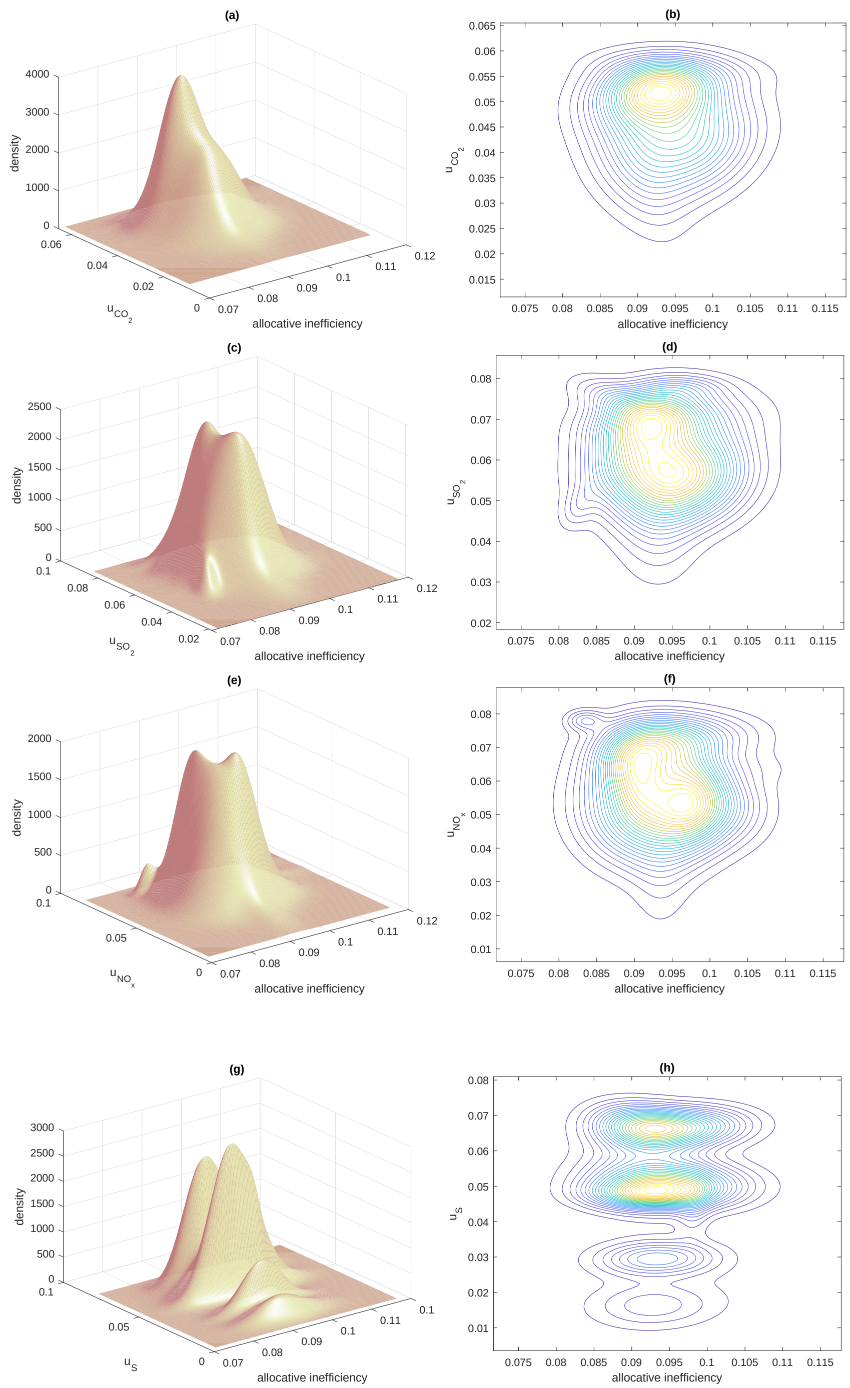


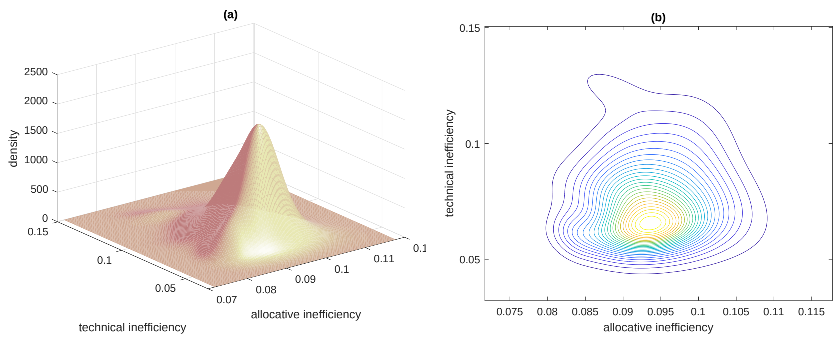
Figure A8 provides the bivariate posterior relationship between AI and the error terms in the MB equations. The interpretation is higly similar to that in Figure A7.

Figure A8: Aspects of bivariate posteriors of allocative inefficiency and one-sided error terms in the MB equations



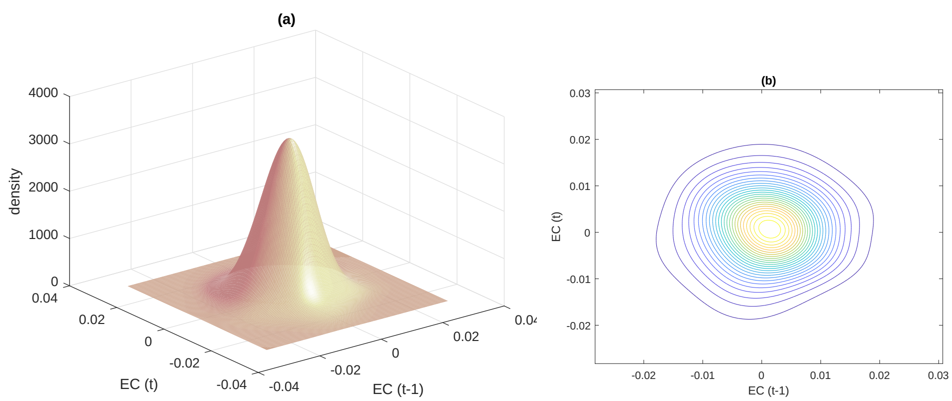
In Figure A9 we present the bivariate posterior for TI and AI. Although this density is uni-modal, there is still considerable opportunity for plants to reduce both types of inefficiency by either adopting better management practices or better technology.

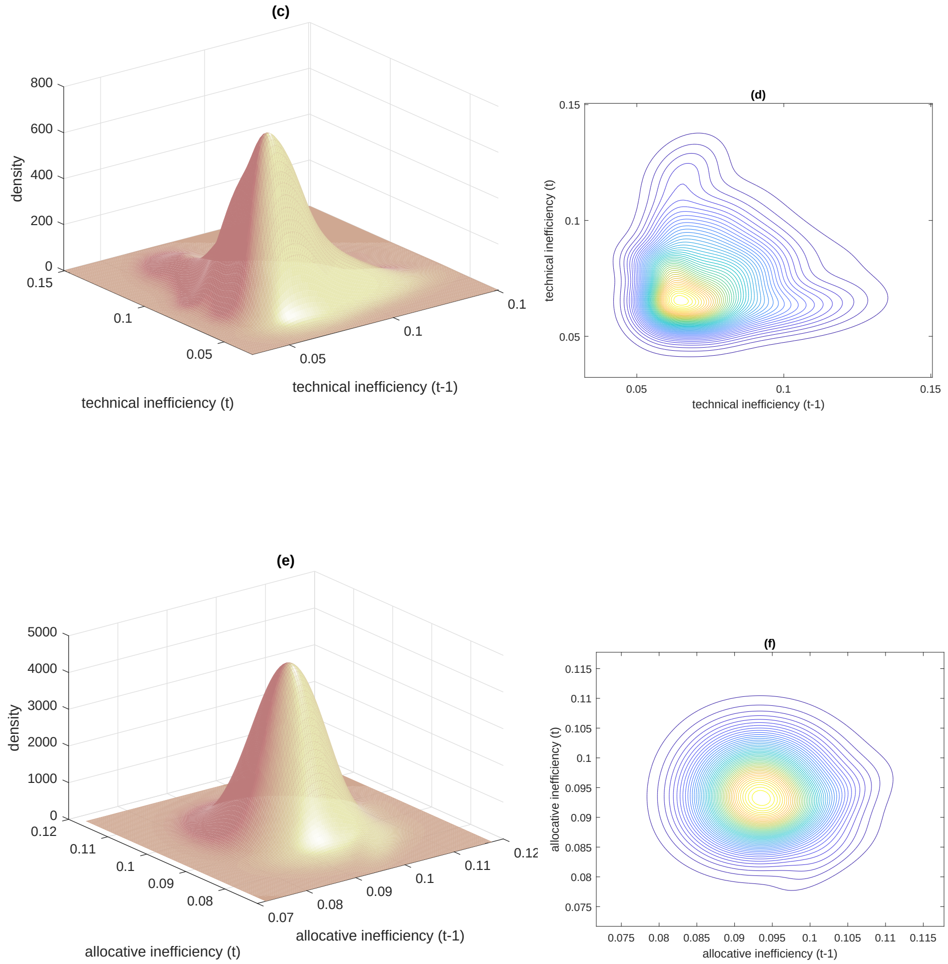
Figure A9: Bivariate posterior of technical and allocative inefficiency



Finally, in Figure A10, we examine the bivariate posteriors for EC, TI, and AI at dates t and t-1. Although these distributions are unimodal, for all three measures there is ample opportunity to improve current-period measures for a given level of performance in the previous period. That is, plants are not locked into poor performance from the previous year.

Figure A10: Bivariate posteriors of EC, TI, and AI at dates t and $t-1$




(c)
density
800
600
400
200
0
0.15
0.1
0.05
technical inefficiency (t)
0.05
0.1
0.1
technical inefficiency (t-1)
(d)
technical inefficiency (t)
0.15
0.1
0.05
0.05
0.1
0.15
technical inefficiency (t-1)
(e)
density
5000
4000
3000
2000
1000
0
0.12
0.11
0.1
0.09
0.08
allocative inefficiency (t)
0.07
0.08
0.09
0.1
0.11
0.12
allocative inefficiency (t-1)
(f)
allocative inefficiency (t)
0.115
0.11
0.105
0.1
0.095
0.09
0.085
0.08
0.075
0.075 0.08 0.085 0.09 0.095 0.1 0.105 0.11 0.115
allocative inefficiency (t-1)