SHADOW DIRECTIONAL DISTANCE FUNCTIONS WITH BADS: GMM ESTIMATION OF OPTIMAL DIRECTIONS AND EFFICIENCIES

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
email: m.tsionas@lancaster.ac.uk

February 16, 2017

*Corresponding author with phone: 706 549 0977; fax: 706 542 3376

Abstract

Because of its greater flexibility, the directional distance function (DDF) has been employed with increasing frequency to estimate multiple-input and multiple-output production, where inputs and outputs can be good or bad. However, typically researchers make three restrictive assumptions. First, they assume a direction of movement of firm production toward the frontier. Second, they assume that actual quantities of inputs and outputs are allocatively or price efficient. Third, they assume exogeneity of all inputs and all outputs, except for the normalized one. The first contribution of this paper is to include parameters to estimate optimal directions which correspond to the firm's profit-maximizing (PM) position. The second contribution is to generalize the DDF to a shadow-quantity DDF. This entails adding distortion parameters to each input and output quantity of the DDF, creating shadow quantities. To estimate the shadow quantities and the structural parameters, we form the shadow DDF system, which includes the shadow DDF and all the first-order price equations from the shadow-PM problem. These include prices for bad inputs and bad outputs, where we approximate their missing prices for use in their first-order price equations. The third contribution is that we estimate the shadow DDF system using a Generalized Method of Moments approach, where all variables are potentially endogenous. This approach is simpler than the Bayesian one employed in Atkinson, Primont, and Tsionas (2016), which estimated shadow prices and optimal directions. Using the same data set, both sets of results are qualitatively very similar, although they differ somewhat quantitatively.

JEL CODES: C11, C33, D24

KEY WORDS: GMM estimation, directional distance function, productivity change with goods and bads, endogeneity, optimal directions.

1 Introduction

The researcher estimating a production technology typically calculates the estimated distance from the production frontier in different time periods. Essential questions are how to treat endogeneity, select the direction for computing the distance, and measure price or allocative inefficiency.[1] Traditional estimation of production technologies using the input- or output-oriented radial distance function (DF) is complicated when inputs or outputs are bads (which are of no utility to the firm).[2] Since the DF scales all good inputs (using the input-oriented DF) or all good outputs (using the output-oriented DF) by the same factor, differential credit is not given for reducing specific inputs or increasing specific outputs. That is, the output-oriented DF does not credit the firm for increasing the production of goods and reducing the production of bads. Instead, one rescales bad outputs and good outputs by the same factor, effectively giving the firm the same credit for increasing bads as increasing goods, which is unreasonable. Thus, many authors have estimated an output-oriented DF and treated bad outputs like good inputs (holding both constant). However, this does not allow crediting the firm for simultaneously reducing good inputs and bad outputs while increasing good outputs. See Atkinson, Primont, and Tsionas (hereafter, APT) (2016) for further discussion. In addition this treatment is inconsistent with the materials-balance principle, which states that the weight of all material outputs of any production process equals the weight of all material inputs. See Ayres and Kneese (1969) for an early discussion of this point. Also see the important recent work by Førsund (2009) and Murty, Russell, and Levkoff (2012). Both papers impose a materials-balance constraint on production function modelling.[3]

More recently, researchers have estimated the directional distance function (DDF) as a less restrictive alternative to the DF. Chambers (1998) and Chambers et al. (1998)

[1]The distance from the frontier is used to compute technical efficiency (TE) in each period as the percent of the efficiency of the frontier firm. Differences betweens distance measures from one period to the next are used to compute productivity change (PC), which is the sum of technical change (TC) (the outward shift in the frontier) and efficiency change (EC) (the extent to which the firm catches up to the frontier). Allocative efficiency means that ratios of prices equal ratios of marginal products for all inputs while price efficiency means that, in addition, each output's price equals its marginal cost.

[2]The DF is input- (output-) oriented if all inputs (outputs) are proportionally scaled down (up) to reach the production frontier while all outputs (inputs) are held constant.

[3]We do not attempt to incorporate a materials-balance approach into this paper but rather leave this as an important topic for future research.

developed the DDF to allow for unique additive changes in each input and output by employing different directions of movement for each to reach the production frontier.[4] For a summary of the theory and application of the DDF see Färe and Grosskopf (2000).[5]

However, researchers typically use restrictive assumptions with DDF estimation. The first of these is to specify arbitrary directions of movement of firm production toward the frontier to measure TE (technical efficiency). As discussed in APT (2016) in greater detail, a number of Data Envelopment Analysis (DEA) studies have attempted to avoid the arbitrary choice of directions by choosing "optimal" directions using linear programming methods to maximize measured TE. These include Färe et al. (2013), Hampf and Krüger (2015), and Zofio et al. (2013). However, by maximizing the measured TE of the firm relative to a DDF, they do not satisfy the first-order conditions for profit maximization (PM), so that their computed directions do not necessarily correspond to optimal directions that satisfy PM conditions.

The first contribution of this paper is to compute optimal directions which move the firm from its current position to a position of PM on the production frontier. We term these directions "optimal-PM" directions.

Typically, productivity studies also make the restrictive assumption that actual quantities of inputs and outputs are price efficient. Thus, the second contribution of this paper is to generalize the DDF to a shadow DDF. This entails adding distortion parameters to each input and output quantity of the DDF, creating *shadow quantities*. These quantities are price efficient subject to profit maximization (PM) based on market prices and allow estimation of input and output inefficiency by comparison with actual quantities. To jointly estimate the shadow quantities, the optimal directions, and the structural parameters, we form the shadow DDF system, which includes the shadow DDF and all of the first-order price equations from the PM problem. Since the prices for bad inputs and bad outputs are missing, we approximate these prices and then use them in their first-order equations. We differ from Atkinson and Tsionas (hereafter, AT) (2016) who

[4] If non-zero directions change only inputs (outputs) when measuring productivity growth, the DDF is input- (output-) oriented. When non-zero directions change all inputs and outputs, the DDF is technology-oriented.

[5] Their joint production approach, discussed later in this paper, and the materials-balance approach of Murty, Russell, and Levkoff (2012) are two alternative ways to model bad outputs. The appropriateness of one or the other would depend on the context and the objective of the researcher.

omit distortion parameters and include price equations only for good inputs and good outputs, since the prices of bad inputs and bad outputs are incomplete. We also differ from APT (2016) who estimated *shadow prices* rather than *shadow quantities*.

Productivity studies also typically assume that inputs and outputs except for the normalized left-hand-side one are exogenous. Our third contribution is to form moment conditions, which we estimate using the Generalized Method of Moments (GMM), where we assume that all variables of the DDF are potentially endogenous. We employ exogenous prices of good inputs and good outputs as part of our instrument set. Our approach is a simpler alternative to Bayesian likelihood-based methods utilized by APT (2016) and AT (2016). We also report bootstrapped finite-sample distributions of distortion parameters and find highly non-normal distributions which have generally little overlap with zero. The shadow DDF system also allows the identification and estimation of measures of TE, PC, TC, and EC.

Our identification strategy relies in part on the use of good input and good output prices as instruments. We argue that good input prices are exogenous, since utilities shop for labor, capital, and energy in competitive national markets. We also argue that good output prices are exogenous. A few of our sample utilities were restructured later in our sample period, so that they sold output in competitive markets. The majority of our sample firms faced regulatory commissions which specified their good output prices.

For an unbalanced panel of U.S. privately-owned utilities, we compute prices of bad inputs and bad outputs, optimal-PM directions, distortion parameters, and the resource implications of price inefficiency. These results are qualitatively similar to those in APT (2016), who estimated optimal-PM directions and *shadow prices* using more complex Bayesian methods with the same data set.

2 The Directional Distance Function

Consider a firm production technology where firms combine 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, $\mathcal{T}(t)$, can be written as

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

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. Following Chambers (1998), we define the technology DDF as

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

where we estimate optimal-PM directions (which can be positive or negative) rather than assigning them a priori, while $\beta \in \mathbb{R}_+$.

Among the important properties of the technology DDF that hold whether directions are assigned or estimated optimally are:

P1. Translation Property: Regardless of the signs of the elements of $\boldsymbol{g}$,

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

as proved in Hudgins and Primont (2007),

P2. $\boldsymbol{g}$-Homogeneity of Degree Minus One:

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

P3. Concavity:

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

P4. Non-negativity:

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

For the four following monotonicity conditions, "S" indicates that the condition depends on the assumption of strong disposability.

P5-S. Good Input Monotonicity: APT (2016) prove, given strong disposability, that

$$\mathbf{x}' \geqq \mathbf{x} \to \overrightarrow{D}_{\mathcal{T}}(\mathbf{x}', \tilde{\mathbf{x}}, \mathbf{y}, \tilde{\mathbf{y}}; \boldsymbol{g}) \geqq \overrightarrow{D}_{\mathcal{T}}(\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y}, \tilde{\mathbf{y}}; \boldsymbol{g}).$$

This follows a similar proof in Chambers (1998) for a benefit function.

Again following APT (2016), we convert this into a partial derivative obtaining

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

P6-S. Good Output Monotonicity: In a similar manner, assuming the strong disposability of $\mathbf{y}$ we can follow P5-S and establish that

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

Likewise, if we make the dubious assumption that bad inputs and bad outputs are strongly disposable, we can follow P5-S and establish P7-S and P8-S.

P7-S. Bad Input Monotonicity:

$$\partial \overrightarrow{D}_{\mathcal{T}}(\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y}, \tilde{\mathbf{y}}; \boldsymbol{g})/\partial \tilde{x}_j \geq 0, j = 1, \ldots, J, \tag{9}$$

P8-S. Bad Output Monotonicity:

$$\partial \overrightarrow{D}_{\mathcal{T}}(\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y}, \tilde{\mathbf{y}}; \boldsymbol{g})/\partial \tilde{y}_l \geq 0, l = 1, \ldots, L. \tag{10}$$

The properties P1-P4 and P5-S–P8-S do not imply that movement from the firm's current position to the PM position on the production frontier requires that bads decrease and goods increase. Färe and Grosskopf (2000) specified these directions of movement a priori before measuring the distance to the frontier. However, they did not move the firm to a PM position. As shown in Fig. 1 of APT (2016), movement of the firm to the frontier at a PM position can result in the increase of both a bad and good output. For example, optimal production could require the simultaneous generation of more electricity and more pollution as a result. We illustrate this in Fig. 1, where the firm could move from an interior point, z, to the optimal profit-maximizing position at z^* or at z', where the price line is tangent to the production possibilities curve. The first move implies that good outputs increase while bad outputs decrease. The second move implies that both good outputs and bad outputs increase.

In addition, we argue that strong disposability and hence P7-S and P8-S would rarely hold true for the electric utility industry. This is because bad inputs and bad outputs are weakly, rather than strongly, disposable for electric utilities. Weak disposability implies that bad inputs and bad outputs cannot be unilaterally decreased, but must be decreased jointly with some other input or output. Bad inputs can be reduced by using fewer good inputs, with which they are bonded chemically. Bad outputs can be reduced by diminishing good outputs and diverting some of the unused inputs to pollution control.

Without the assumption of strong disposability, we are unable to derive general monotonicity conditions for bad inputs and bad outputs. However, if we assume that their prices are positive and assume that firms maximize profits, we can determine *local* monotonicity conditions for bad inputs and bad outputs.

Many studies have maintained the assumption that the electric utility industry maximizes profits. One of the earliest was Atkinson and Halvorsen (1976). More recently see AT (2016). Fowlie (2010) provides evidence of unregulated profit-maximizing behavior by finding that many regulated utilities earn allowed rates of return on capital that considerably exceed the market rate of return (implying that constraints on PM are not binding). A more recent example is APT (2016), which compares models based on PM and cost minimization. Using Bayesian criteria, they find that the former model performs considerably better.

Following Chambers et al. (1998) and Färe and Grosskopf (2000), we assume that firms maximize profits by choosing optimal values for $\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y}, \tilde{\mathbf{y}}$:

$$\sup \big\{ \mathbf{p}_y(\mathbf{y} + \overrightarrow{D}_{\mathcal{T}} \boldsymbol{g}_y) - \mathbf{p}_{\tilde{y}}(\tilde{\mathbf{y}} + \overrightarrow{D}_{\mathcal{T}} \boldsymbol{g}_{\tilde{y}}) - \mathbf{p}_x(\mathbf{x} + \overrightarrow{D}_{\mathcal{T}} \boldsymbol{g}_x) - \mathbf{p}_{\tilde{x}}(\tilde{\mathbf{x}} + \overrightarrow{D}_{\mathcal{T}} \boldsymbol{g}_{\tilde{x}}) \big\}, \qquad (11)$$

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. As discussed in more detail in APT (2016), all prices are assumed to be non-negative except for the price of bad inputs, which is non-positive. The price of a bad input is non-positive, since the seller must reduce the price to the firm as compensation for consuming more of the bad, all else constant. The price of a regulated bad output subject to a binding emission standard is positive. The firm must buy permits for any emissions exceeding what are allowed under the emission standard set by the states (which in turn must meet Federal air quality standards). Each year under the cap-and-trade system for SO_2, allowable emissions were decreased, so that all the utilities in our sample faced a binding emission standard during our sample period. The limited number of published permit prices for SO_2 were positive and quite high. Failure to purchase SO_2 permits to cover excess emissions resulted in a very large fine.

We have only highly limited price series for bad outputs and bad inputs. Prices of SO_2 and NO_x pollution permits are unavailable for most utilities in most of our sample time periods. Markets for these permits are very thin. In addition, we are unable to compute an hedonic price for sulfur. We do have data on the delivered prices of coal and oil. However, these prices (per ton and per barrel) are a function of the Btu content, the sulfur content, and transportation costs. Since the latter are confidential, we cannot run a hedonic regression to compute the implicit price of sulfur. Data for mine-mouth prices of coal and well-head prices of oil would be free of transportation charges. If these prices were available, one then could regress them on sulfur and Btu content to obtain the implicit price of sulfur. However, mine-mouth and well-head price data are also confidential.

The first-order PM conditions for inputs and outputs are given in APT (2016) as:

$$
\begin{aligned}
p_n/\vartheta &= \partial\overrightarrow{D}_{\mathcal{T}}(\mathbf{x},\tilde{\mathbf{x}},\mathbf{y},\tilde{\mathbf{y}};\boldsymbol{g})/\partial x_n, n=1,\ldots,N, && (12)\\
p_m/\vartheta &= -\partial\overrightarrow{D}_{\mathcal{T}}(\mathbf{x},\tilde{\mathbf{x}},\mathbf{y},\tilde{\mathbf{y}};\boldsymbol{g})/\partial y_m, m=1,\ldots,M, && (13)\\
p_j/\vartheta &= \partial\overrightarrow{D}_{\mathcal{T}}(\mathbf{x},\tilde{\mathbf{x}},\mathbf{y},\tilde{\mathbf{y}};\boldsymbol{g})/\partial \tilde{x}_j, j=1,\ldots,J, && (14)\\
p_l/\vartheta &= \partial\overrightarrow{D}_{\mathcal{T}}(\mathbf{x},\tilde{\mathbf{x}},\mathbf{y},\tilde{\mathbf{y}};\boldsymbol{g})/\partial \tilde{y}_l, l=1,\ldots,L, && (15)
\end{aligned}
$$

where

$$\vartheta = [\sum_m p_m g_m - \sum_n p_n g_n - \sum_j p_j g_j - \sum_l p_l g_l]. \tag{16}$$

We assume that $\vartheta > 0$, where ϑ is the optimal value of the Lagrangian multiplier. which is the change in profits due to a small improvement in the production technology. For details see Hudgins and Primont (2007) who show that one can solve the unconstrained profit-maximization problem in (11) or solve the equivalent Lagrangian function as

$$\mathcal{L} = \mathbf{p}_y\mathbf{y} - \mathbf{p}_x\mathbf{x} + \vartheta\overrightarrow{D}_{\mathcal{T}}(\mathbf{x},\tilde{\mathbf{x}},\mathbf{y},\tilde{\mathbf{y}};\boldsymbol{g}), \tag{17}$$

in order to obtain this interpretation of ϑ.

Then we can state the monotonicity conditions for the case of weakly disposable inputs and outputs, where "W" indicates weakly disposable. From (14) and the assumption that $\mathbf{p}_{\tilde{x}} \leq 0$ we obtain locally

P7-W. Bad Input Monotonicity:

$$\partial\overrightarrow{D}_{\mathcal{T}}(\mathbf{x},\tilde{\mathbf{x}},\mathbf{y},\tilde{\mathbf{y}};\boldsymbol{g})/\partial \tilde{x}_j \leq 0, j=1,\ldots,J. \tag{18}$$

From (15) and the assumption that $\mathbf{p}_{\tilde{y}} \geq 0$ we obtain locally

P8-W. Bad Output Monotonicity:

$$\partial \overrightarrow{D}_{\mathcal{T}}(\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y}, \tilde{\mathbf{y}}; \boldsymbol{g})/\partial \tilde{y}_l \geq 0, l = 1, \ldots, L. \tag{19}$$

3 The Econometric Model

3.1 The Shadow Distance System

To simplify notation, let $\mathbf{z} = (\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y}, \tilde{\mathbf{y}})$. Assume that we have panel data for firm $i (i = 1, \ldots, F)$ in time period $t (t = 1, \ldots, T)$ on all inputs and outputs. We add vintage (τ) (which is an output-weighted-average of the age of the firm's capital) to account for the fact that capital depreciates with age. We employ a quadratic function of all inputs, outputs, τ, time, and distortion parameters, $\boldsymbol{k} = (k_n, k_m, k_j, k_l)$, which measure deviations of shadow (efficient) quantities from actual quantities. This provides a new formulation of the technology DDF, which we term a shadow technology DDF:

$$\begin{aligned}
0 &= \overrightarrow{D}_{\mathcal{T}}(\mathbf{z}, \tau, t; \boldsymbol{k}) + \epsilon_{it} \\
&= \sum_{m=1}^{M} \gamma_m (y_{m,it} + k_m) + \sum_{l=1}^{L} \gamma_l (\tilde{y}_{l,it} + k_l) \\
&+ \sum_{n=1}^{N} \gamma_n (x_{n,it} + k_n) + \sum_{j=1}^{J} \gamma_j (\tilde{x}_{j,it} + k_j) \\
&+ \frac{1}{2} \sum_{m=1}^{M} \sum_{m'=1}^{M} \gamma_{mm'} (y_{m,it} + k_m)(y_{m',it} + k_{m'}) + \frac{1}{2} \sum_{l=1}^{L} \sum_{l'=1}^{L} \gamma_{ll'} (\tilde{y}_{l,it} + k_l)(\tilde{y}_{l',it} + k_{l'}) \\
&+ \frac{1}{2} \sum_{n=1}^{N} \sum_{n'=1}^{N} \gamma_{nn'} (x_{n,it} + k_n)(x_{n',it} + k_{n'}) + \frac{1}{2} \sum_{j=1}^{J} \sum_{j'=1}^{J} \gamma_{jj'} (\tilde{x}_{j,it} + k_j)(\tilde{x}_{j',it} + k_{j'}) \\
&+ \sum_{j=1}^{J} \sum_{n=1}^{N} \gamma_{jn} (\tilde{x}_{j,it} + k_j)(x_{n,it} + k_n) + \sum_{m=1}^{M} \sum_{n=1}^{N} \gamma_{mn} (y_{m,it} + k_m)(x_{n,it} + k_n) \\
&+ \sum_{l=1}^{L} \sum_{n=1}^{N} \gamma_{ln} (\tilde{y}_{l,it} + k_l)(x_{n,it} + k_n) + \sum_{m=1}^{M} \sum_{j=1}^{J} \gamma_{jm} (y_{m,it} + k_m)(\tilde{x}_{j,it} + k_j) \\
&+ \sum_{l=1}^{L} \sum_{j=1}^{J} \gamma_{jl} (\tilde{y}_{l,it} + k_l)(\tilde{x}_{j,it} + k_j) + \sum_{l=1}^{L} \sum_{m=1}^{M} \gamma_{lm} (\tilde{y}_{l,it} + k_l)(y_{m,it} + k_m)
\end{aligned}$$

$$
\begin{aligned}
&+ \sum_{m=1}^{M} \gamma_{mt}(y_{m,it} + k_m)\, t + \sum_{l=1}^{L} \gamma_{lt}(\tilde{y}_{l,it} + k_l)\, t + \sum_{n=1}^{N} \gamma_{nt}(x_{n,it} + k_n)\, t \\
&+ \sum_{j=1}^{J} \gamma_{jt}(\tilde{x}_{j,it} + k_j)\, t + \sum_{t} \gamma_t d_t + \sum_{i} d_i + \gamma_\tau \tau_{it} + \epsilon_{it},
\end{aligned} \tag{20}
$$

where

$$
\epsilon_{it} = v_{it} - u_{it}, \tag{21}
$$

so that ϵ_{it} is an additive error with a one-sided component, u_{it}, and a standard noise component, v_{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 firm dummy, and the $k_q, q = n, m, j, l$ are parameters which define shadow quantities as $x_q + k_q, q = n, m, j, l$. APT (2016) provides the restrictions on the parameters of the technology distance function ($\boldsymbol{\gamma}$) and the optimal-PM directions ($\boldsymbol{g}$) that impose the translation property in (3) for a technology-oriented DDF:

$$
\begin{aligned}
\sum_{m=1}^{M} \gamma_m g_m + \sum_{l=1}^{L} \gamma_l g_l + \sum_{n=1}^{N} \gamma_n g_n + \sum_{j=1}^{J} \gamma_j g_j &= -1, \\
\sum_{m=1}^{M} \gamma_{mn'} g_m + \sum_{l=1}^{L} \gamma_{ln'} g_l + \sum_{n=1}^{N} \gamma_{nn'} g_n + \sum_{j=1}^{J} \gamma_{jn'} g_j &= 0, \ \forall\, n', \\
\sum_{m=1}^{M} \gamma_{mm'} g_m + \sum_{l=1}^{L} \gamma_{lm'} g_l + \sum_{n=1}^{N} \gamma_{m'n} g_n + \sum_{j=1}^{J} \gamma_{jm'} g_j &= 0, \ \forall\, m', \\
\sum_{m=1}^{M} \gamma_{j'm} g_m + \sum_{l=1}^{L} \gamma_{j'l} g_l + \sum_{n=1}^{N} \gamma_{j'n} g_n + \sum_{j=1}^{J} \gamma_{jj'} g_j &= 0, \ \forall\, j', \\
\sum_{m=1}^{M} \gamma_{l'm} g_m + \sum_{l=1}^{L} \gamma_{ll'} g_l + \sum_{n=1}^{N} \gamma_{l'n} g_n + \sum_{j=1}^{J} \gamma_{jl'} g_j &= 0, \ \forall\, l'.
\end{aligned} \tag{22}
$$

We impose the translation property restrictions in (22) on (20) to yield $\overrightarrow{D}_{\mathcal{T}}(\mathbf{z}, \tau, t; \boldsymbol{k}, \boldsymbol{g})$. Note that parameters which measure optimal-PM directions ($\boldsymbol{g}$) occur only in these constraints. Let us rewrite the above restrictions more compactly in the following form:

$$
\mathbf{h}\,(\boldsymbol{\theta}) = \mathbf{c}, \tag{23}
$$

where $\mathbf{c} = [-1\vdots\mathbf{0}]$ and $\boldsymbol{\theta} = [\boldsymbol{\gamma}', \boldsymbol{g}']$.

We can rewrite the first-order PM conditions from (12)-(15) in terms of the parameters of the quadratic shadow DDF in (20) for each good input price equation as:

$$p_{n,it}/\vartheta = \gamma_n + \sum_{n'=1}^{N} \gamma_{nn'}(x_{n',it} + k_{n'}) + \sum_{j=1}^{J} \gamma_{jn}(\tilde{x}_{j,it} + k_j) + \sum_{m=1}^{M} \gamma_{mn}(y_{m,it} + k_m) + \sum_{l=1}^{L} \gamma_{ln}(\tilde{y}_{l,it} + k_l) + \gamma_{nt}t + v_{n,it}, \quad (24)$$

for each good output price equation as:

$$p_m/\vartheta = -\left[\gamma_m + \sum_{m'=1}^{M} \gamma_{mm'}(y_{m',it} + k_{m'}) + \sum_{j=1}^{J} \gamma_{jm}(\tilde{x}_{j,it} + k_j) + \sum_{n=1}^{N} \gamma_{mn}(x_{n,it} + k_n) + \sum_{l=1}^{L} \gamma_{lm}(\tilde{y}_{l,it} + k_l) + \gamma_{mt}t\right] + v_{m,it}, (25)$$

for each bad input price equation as:

$$p_j/\vartheta = \gamma_j + \sum_{j'=1}^{J} \gamma_{jj'}(\tilde{x}_{j',it} + k_{j'}) + \sum_{n=1}^{N} \gamma_{jn}(x_{n,it} + k_n) + \sum_{m=1}^{M} \gamma_{jm}(y_{m,it} + k_m) + \sum_{l=1}^{L} \gamma_{jl}(\tilde{y}_{l,it} + k_l) + \gamma_{jt}t + v_{j,it}, \quad (26)$$

and for each bad output price equation as:

$$p_l/\vartheta = \gamma_l + \sum_{l'=1}^{L} \gamma_{ll'}(\tilde{y}_{l',it} + k_{l'}) + \sum_{j=1}^{J} \gamma_{jl}(\tilde{x}_{j,it} + k_j) + \sum_{n=1}^{N} \gamma_{ln}(x_{n,it} + k_n) + \sum_{m=1}^{M} \gamma_{lm}(y_{m,it} + k_m) + \gamma_{\tilde{y}t}t + v_{l,it}, \quad (27)$$

where the $v_{g,it},\ g = n, m, j, l$ have zero mean. We impose the restrictions in (22) on (24)-(27).

We do not have prices for bad inputs or bad outputs. However, we include the first-order conditions involving latent prices for bad outputs and bad inputs with equations

(26) and (27). Their latent prices are assumed to follow the process:

$$p_{j,it}/\vartheta_{it} = \sum_{t=1}^{T} \eta_{ijt}^{(1)} d_t + \sum_{i=1}^{F} \eta_{ijt}^{(2)} d_i + \eta_{ijt}^{(3)} \tau_{it} + \eta_{ijt}^{(4)} p_{E,it}, \ \forall j = 1, \ldots, J, \tag{28}$$

and

$$p_{l,it}/\vartheta_{it} = \sum_{t=1}^{T} \eta_{ilt}^{(1)} d_t + \sum_{i=1}^{F} \eta_{ilt}^{(2)} d_i + \eta_{ilt}^{(3)} \tau_{it} + \eta_{ilt}^{(4)} p_{E,it}, \ \forall l = 1, \ldots, L, \tag{29}$$

where $\{d_t, t = 1, ..., T; i = 1, ..., F\}$ is a full set of time dummies, $\{d_i, i = 1, ..., F; t = 1, ..., T\}$ is a full set of firm dummies, $p_{E,it}$ is the price of energy relative to capital, and $\eta_{ijt}^{(1)}, \eta_{ijt}^{(2)}$, $\eta_{ilt}^{(1)}, \eta_{ilt}^{(2)}$ are firm-specific coefficients with the normalization $\sum_{t=1}^{T} \eta_{ijt}^{(1)} = \sum_{i=1}^{F} \eta_{ijt}^{(2)} = \sum_{t=1}^{T} \eta_{ilt}^{(1)} = \sum_{i=1}^{F} \eta_{ilt}^{(2)} = 0$. Specifically, we rewrite each constraint so that the first firm (among the F) and the first time period (among the T) appear on the left-hand-side of each constraint equation before substitution into (28) and (29).

The firm-specific η_{ijt} and η_{ilt} in (28) and (29) are still not yet identifiable. To identify them we assume that:

$$\eta_{ijt}^{(\varrho)} = \sum_{i=1}^{F} \delta_{ij1}^{(\varrho)} d_i + \delta_{ij2}^{(\varrho)} \tau_{it} + \delta_{ij3}^{(\varrho)} p_{E,it}, \ \varrho = 1, ..., 4; j = 1, ..., J. \tag{30}$$

$$\eta_{ilt}^{(\varrho)} = \sum_{i=1}^{F} \delta_{il1}^{(\varrho)} d_i + \delta_{il2}^{(\varrho)} \tau_{it} + \delta_{il3}^{(\varrho)} p_{E,it}, \ \varrho = 1, ..., 4; l = 1, ..., L. \tag{31}$$

We then substitute (30) and (31) into (28) and (29). Then these latter two equations replace $p_{j,it}/\vartheta_{it}$ and $p_{l,it}/\vartheta_{it}$ in the LHS of (26) and (27). The resulting equations do not involve non-linearities in the parameters, but instead, non-linearities in the variables, so that estimation is straightforward. Non-linearities in the variables are necessary to identify the firm-specific parameters. Specifically, identification results from assuming that the η_{ijt} and η_{ilt} depend on d_i, τ_{it}, and $p_{E,it}$, all of which involve the subscript i, and the non-linearities obtained through substitution of (30) and (31) into (28) and (29). There is no empirical complication to replacing $p_{j,it}/\vartheta_{it}$ and $p_{l,it}/\vartheta_{it}$ with a function of additional parameters and variables, since in Section 4 we carry out GMM estimation. This simply requires rewriting the set of estimated equations, (24)-(27) and (20) in terms of their error terms on the left-hand-side and then constructing the sample moment conditions in (46). Some or all of the terms in (30) and (31) could be employed in estimation. Below we report results that indicate the superiority of the full specification of these equations. Clearly, if one does not wish to obtain firm-specific parameters, then (30) and (31) would not be employed. This approach to generating latent prices is much simpler than the Bayesian approach utilized in APT (2016).

3.2 Shadow Prices and Shadow Quantities

Two methods exist to estimate the extent of resource misallocation. APT (2016) estimated a DDF system assuming PM and incorporated additive terms in the first-order price equations to measure deviations of shadow prices from market prices. Firms are assumed to determine actual levels of input and output quantities using shadow prices (prices that are relevant to the firm). Actual quantities are inefficient if shadow prices differ from actual (market) prices. Shadow prices may deviate from actual prices due to taxes, subsidies, rate-of-return regulation on capital, and other similar reasons. Then APT (2016) calculated the inefficient usage of inputs and generation of outputs from the fitted first-order price equations. This involved setting shadow prices equal to market prices and computing the implied shadow quantities.

In this paper, instead of estimating shadow prices that are consistent with actual quantities as in APT (2016), we estimate shadow quantities that are consistent with actual prices. This requires generalizing the DDF to a shadow DDF, where shadow quantities replace actual ones. Resources are misallocated by the amount that directly computed shadow quantities (based on actual prices) differ from actual quantities. That is, if actual values of $x_q, q = n, m, j, l$ can be adjusted up or down by a non-zero k_q to increase price efficiency, then actual quantities are inefficient by that amount. We use bootstrap methods to compute distributions for parameters that measure inefficient resource allocation.

3.3 Measurement of Inefficiencies

We assume that u_{it} is a one-sided, non-negative error term in (21). Following Cornwell, Schmidt, and Sickles (1990) we regress the residual $\hat{\epsilon}_{it}$ on a function of time and firm dummies:

$$\hat{\epsilon}_{it} = \sum_i f_i d_i + b_i d_i t + c_i d_i t^2 + \xi_{it}, \tag{32}$$

where d_i are firm dummies (and coefficients b_i and c_i result from interacting trend and trend squared with the dummies in d_i) while ξ_{it} is a two-sided i.i.d. error term. The fitted values of this equation are $\hat{u}_{it}$. Then, we compute TE_{it} as $\exp(-\hat{u}_{it})$ and report its distribution for our sample.

Estimation of price efficiency requires identification of the k_q distortion parameters. For all inputs and outputs, we first specify equations for input-specific and output-specific (constant across firms) k values, (k_n, k_m, k_j, k_l), as

$$k_n = \sum_t k_{nt} d_t, \ \forall \ n = 1, \ldots, N-1, \tag{33}$$

$$k_m = \sum_t k_{mt} d_t, \ \forall \ m = 1, \ldots, M, \tag{34}$$

$$k_j = \sum_t k_{jt} d_t, \ \forall \ j = 1, \ldots, J, \tag{35}$$

$$k_l = \sum_t k_{lt} d_t, \ \forall \ l = 1, \ldots, L, \tag{36}$$

We refer to these as Case A.

We also specify equations for input- and firm-specific k values, $(k_{n,i}, k_{m,i}, k_{j,i}, k_{l,i})$, as

$$k_{n,i} = \sum_t k_{nt} d_t + \sum_t k_{n,it} d_i d_t, \ \forall\ n = 1, \ldots, N-1; i = 1, \ldots, F, \tag{37}$$

$$k_{m,i} = \sum_t k_{mt} d_t + \sum_t k_{m,it} d_i d_t, \ \forall\ m = 1, \ldots, M; i = 1, \ldots, F, \tag{38}$$

$$k_{j,i} = \sum_t k_{jt} d_t + \sum_t k_{j,it} d_i d_t, \ \forall\ j = 1, \ldots, J; i = 1, \ldots, F, \tag{39}$$

$$k_{l,i} = \sum_t k_{lt} d_t + \sum_t k_{l,it} d_i d_t, \ \forall\ l = 1, \ldots, L; i = 1, \ldots, F. \tag{40}$$

We refer to these as Case B and use these equations to replace the input-specific k values in (20).

This approach assumes the errors in allocative efficiency can be specified as parameters, following the work of Atkinson and Halvorsen (1976), Atkinson and Primont (2002), and Kumbhakar (1992). A random effects approach is an alternative as in Kumbhakar and Tsionas (2005). The current approach avoids the need to make distributional assumptions and simplifies the estimation process relative to the random effects approach.

Standardized data are employed so that comparisons of efficiency parameters and optimal-PM directions are not dependent on the unit of measure (i.e., the scale of the data). If the k_q are not significantly different from zero, we fail to reject the null that firms are employing the resource q in a price-efficient manner. We compute the percent change in input and output usage, $\%\Delta_q$, required to achieve their price efficient level. We define $\%\Delta_q = [(x_q + k_q) - x_q]/x_q = k_q/x_q$, which is the shadow quantity minus the actual quantity divided by the actual quantity. Thus if k_q is estimated to be negative, the efficient quantity is less than the actual quantity.

3.4 Satisfying P1-P8

As indicated above, we satisfy the translation property of the DDF, P1, by imposing on (20) the translation property restrictions, (22), for the technology-oriented DDF. In APT (2016) we show that P1 for our DDF system implies P2. We have imposed concavity, P3, on our model at random points and found violations in less than 1 % of the observations. After estimation, we impose non-negativity, P4, for all observations via a normalization of the fitted DDF. Finally, our fitted model satisfies the monotonicity properties P5-S, P6-S, p7-W, and P8-W for approximately 99% of the data.

3.5 Estimation of Productivity Change

The calculation of PC and its decomposition into EC and TC proceeds by focusing on (20). As indicated above, in each period we normalize our estimates of $\overrightarrow{D}_{\mathcal{T}}(\mathbf{z}, \tau, t; \boldsymbol{k}, \boldsymbol{g})$ so that all estimated values of $\overrightarrow{D}_{\mathcal{T}}(\mathbf{z}, \tau, t; \boldsymbol{k}, \boldsymbol{g}) \geq 0$ and the frontier firm has an estimated $\overrightarrow{D}_{\mathcal{T}}(\mathbf{z}, \tau, t; \boldsymbol{k}, \boldsymbol{g}) = 0$. Adding a superscript t to indicate the time period (but temporarily suppressing $\tau, \boldsymbol{k}$, and $\boldsymbol{g}$), we eliminate the effect of arbitrarily scaling of the data by defining a percentage change Luenberger technical change measure, $\mathrm{TC}^{\mathcal{L}}$, as

$$\begin{aligned} \mathrm{TC}^{\mathcal{L}}_{it} &= .5\Bigg\{ \frac{\overrightarrow{D}^{t+1}_{\mathcal{T}}(\mathbf{x}_{i,t+1}, \tilde{\mathbf{x}}_{i,t+1}, \mathbf{y}_{i,t+1}, \tilde{\mathbf{y}}_{i,t+1}) - \overrightarrow{D}^{t}_{\mathcal{T}}(\mathbf{x}_{i,t+1}, \tilde{\mathbf{x}}_{i,t+1}, \mathbf{y}_{i,t+1}, \tilde{\mathbf{y}}_{i,t+1})}{A} \\ &\qquad + \frac{\overrightarrow{D}^{t+1}_{\mathcal{T}}(\mathbf{x}_{it}, \tilde{\mathbf{x}}_{it}, \mathbf{y}_{it}, \tilde{\mathbf{y}}_{it}) - \overrightarrow{D}^{t}_{\mathcal{T}}(\mathbf{x}_{it}, \tilde{\mathbf{x}}_{it}, \mathbf{y}_{it}, \tilde{\mathbf{y}}_{it})}{B} \Bigg\}, \end{aligned} \tag{41}$$

where

$$\begin{aligned} A &= .5\{\overrightarrow{D}^{t+1}_{\mathcal{T}}(\mathbf{x}_{i,t+1}, \tilde{\mathbf{x}}_{i,t+1}, \mathbf{y}_{i,t+1}, \tilde{\mathbf{y}}_{i,t+1}) + \overrightarrow{D}^{t}_{\mathcal{T}}(\mathbf{x}_{i,t+1}, \tilde{\mathbf{x}}_{i,t+1}, \mathbf{y}_{i,t+1}, \tilde{\mathbf{y}}_{i,t+1})\} \\ B &= .5\{\overrightarrow{D}^{t+1}_{\mathcal{T}}(\mathbf{x}_{it}, \tilde{\mathbf{x}}_{it}, \mathbf{y}_{it}, \tilde{\mathbf{y}}_{it}) + \overrightarrow{D}^{t}_{\mathcal{T}}(\mathbf{x}_{it}, \tilde{\mathbf{x}}_{it}, \mathbf{y}_{it}, \tilde{\mathbf{y}}_{it})\}. \end{aligned} \tag{42}$$

We define a percentage change Luenberger efficiency change measure, $\mathrm{EC}^{\mathcal{L}}$, as

$$\mathrm{EC}^{\mathcal{L}}_{it} = \frac{\{\overrightarrow{D}^{t}_{\mathcal{T}}(\mathbf{x}_{it}, \tilde{\mathbf{x}}_{it}, \mathbf{y}_{it}, \tilde{\mathbf{y}}_{it}) - \overrightarrow{D}^{t+1}_{\mathcal{T}}(\mathbf{x}_{i,t+1}, \tilde{\mathbf{x}}_{i,t+1}, \mathbf{y}_{i,t+1}, \tilde{\mathbf{y}}_{i,t+1})\}}{C}, \tag{43}$$

where $C = .5\{\overrightarrow{D}^{t}_{\mathcal{T}}(\mathbf{x}_{it}, \tilde{\mathbf{x}}_{it}, \mathbf{y}_{it}, \tilde{\mathbf{y}}_{it}) + \overrightarrow{D}^{t+1}_{\mathcal{T}}(\mathbf{x}_{i,t+1}, \tilde{\mathbf{x}}_{i,t+1}, \mathbf{y}_{i,t+1}, \tilde{\mathbf{y}}_{i,t+1})\}$. Note that the bases–A, B, and C–are midpoints between one fitted DDF and an adjacent one. Finally, the percentage change Luenberger productivity change indicator, $\mathrm{PC}^{\mathcal{L}}_{\mathrm{it}}$, is defined as

$$\mathrm{PC}^{\mathcal{L}}_{it} = \mathrm{TC}^{\mathcal{L}}_{it} + \mathrm{EC}^{\mathcal{L}}_{it}. \tag{44}$$

If the frontier firm is the same in two adjoining periods, the numerators and denominators in (41) and (43) would be zero. In this case, we set $\mathrm{EC}^{\mathcal{L}}_{\mathcal{F}t} = \mathrm{TC}^{\mathcal{L}}_{\mathcal{F}t} = \mathrm{PC}^{\mathcal{L}}_{\mathcal{F}t} = 0$, since the numerator value is zero for the frontier firm, $\mathcal{F}$.

4 GMM-Based Methods

As indicated above, we first impose the translation property restrictions in (22) on the DDF, (20). Then we substitute the price efficiency equations, (37)-(40), into (20), the DDF, to yield $\overrightarrow{D}_{\mathcal{T}}(\mathbf{z}, \tau, t; \boldsymbol{k}, \boldsymbol{g})$ and into (24)-(27), the price equations. Since prices for bad inputs and outputs are missing, we substitute the firm-specific latent price equations, (30) and (31), into the original latent price equations, (28) and (29), and then substitute these resulting equations into (26)-(27). The resulting restricted DDF and restricted price share equations, which are now functions of $(\mathbf{z}, t; \boldsymbol{k}, \boldsymbol{g})$, comprise our shadow DDF system. which allows identification of parameters measuring optimal-PM directions ($\boldsymbol{g}$), the structural parameters of the DDF ($\boldsymbol{\gamma}$), parameters measuring deviations of shadow quantities from actual quantities ($\boldsymbol{k}$), and latent price parameters.

Let us formally specify the moment conditions using the restricted versions of (20) and (24)-(27). First, as indicted above, we must rewrite these equations so that the error of each equation is on its left-hand-side. Then the moment conditions for all of these equations are:

$$\mathbb{E}\left[\mathbf{Z}'_{it}\boldsymbol{V}_{it}(\boldsymbol{\theta})\right] = \mathbf{0}, \tag{45}$$

where $\mathbf{Z}_{it}(i = 1, ..., F; t = 1, \ldots, T)$ denotes a column vector of instruments (exogenous variables including the $\mathbf{X}_{it}$ regressors), such that $\dim(\mathbf{Z}_{it})$ is at least as large as the number of unknown parameters in the system, $\boldsymbol{\theta} = [\boldsymbol{\gamma}', \boldsymbol{g}']' \in \boldsymbol{\Theta} \subseteq \mathbb{R}^p$ is the parameter vector, and $V_{it}(\boldsymbol{\theta}) = [(v_{it} - u_{it})]', \mathbf{v}'_{n,it}, \mathbf{v}'_{m,it}, \mathbf{v}'_{j,it}, \mathbf{v}'_{l,it}]'$, where $\mathbf{v}_{n,it} = [v_{1,it}, \ldots, v_{N,it}]', \mathbf{v}_{m,it} = [v_{1,it}, \ldots, v_{M,it}]', \mathbf{v}_{j,it} = [v_{1,it}, \ldots, v_{J,it}]', \mathbf{v}_{l,it} = [v_{1,it}, \ldots, v_{L,it}]'$.

The sample moment conditions are:

$$\mathbf{G}(\boldsymbol{\theta}; \mathbf{Z}) = (FT)^{-1} \sum_{i,t} \mathbf{Z}'_{it} \boldsymbol{V}_{it}(\boldsymbol{\theta}) = \mathbf{0}. \tag{46}$$

For estimation purposes, we adopt a Generalized Method of Moments-Continuously Updated Estimator (GMM-CUE), which relies on moment conditions generated from (20) and (24)-(27). The modified GMM criterion that we minimize is the following penalized version of GMM:

$$\min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} : \mathbf{G}(\boldsymbol{\theta}; \mathbf{Z})' \boldsymbol{\Omega}(\boldsymbol{\theta})^{-1} \mathbf{G}(\boldsymbol{\theta}; \mathbf{Z}) + (FT)^{-1} \left\{ \frac{1}{\omega_g^2} (\boldsymbol{g} - \tilde{\boldsymbol{g}})' (\boldsymbol{g} - \tilde{\boldsymbol{g}}) + \frac{1}{\omega^2} [\mathbf{h}(\boldsymbol{\theta}) - \mathbf{c}]' [\mathbf{h}(\boldsymbol{\theta}) - \mathbf{c}] \right\}, \tag{47}$$

where $\boldsymbol{\Omega}(\boldsymbol{\theta})$ is the usual GMM weighting matrix used for the CUE version of GMM. Recall that $\boldsymbol{g}$ is a vector of optimal-PM directions and that $\tilde{\boldsymbol{g}} = (\tilde{\boldsymbol{g}}_x, \tilde{\boldsymbol{g}}_{\tilde{x}}, \tilde{\boldsymbol{g}}_y, \tilde{\boldsymbol{g}}_{\tilde{y}}) = (-\mathbf{1}, -\mathbf{1}, +\mathbf{1}, -\mathbf{1})$ is a vector of initial conditions for the directions. These values are typically selected as fixed directions in the literature and hence make reasonable initial values. The required translation property restrictions are imposed through the regularization ("prior") in (23). The weight matrix is given by

$$\boldsymbol{\Omega}(\boldsymbol{\theta}) = (FT)^{-1} \sum_{i,t} \left[\mathbf{G}(\mathbf{Z}_{it}; \boldsymbol{\theta}) \mathbf{G}(\mathbf{Z}_{it}; \boldsymbol{\theta})' \right]. \tag{48}$$

Clearly, we want ω_g to be large and ω to be small. When ω_g is large we make the first half of the penalty function small so that we place a weak constraint on the values that the optimally chosen directions can assume. When ω is small we make the second half of the penalty function large, so that the translation property restrictions will be satisfied with near equality. Given the form of the criterion we see that (47) acts to regularize estimation in the context of GMM. For example, in the normal linear model $\mathbf{y} = \mathbf{Z}\boldsymbol{\beta} + \mathbf{v}$ the prior $\boldsymbol{\beta} \sim \mathscr{N}\left(\mathbf{0}, \omega_g^2 \mathbf{I}\right)$ produces the ridge or regularized estimator $\hat{\boldsymbol{\beta}} = \left(\mathbf{Z}'\mathbf{Z} + \omega_g^{-2} \boldsymbol{I}\right)^{-1} \mathbf{Z}'\mathbf{y}$. The set $\boldsymbol{\Theta}$ guarantees that certain regularity restrictions on the technology directional distance function are imposed at the mean of the data which is normalized to zero for simplicity. In a Bayesian setting our penalty term for $\boldsymbol{g}$ is equivalent to the following:

$$\boldsymbol{g}|\gamma \sim \mathscr{N}\left(\tilde{\boldsymbol{g}}, \omega_g^2 \mathbf{I}\right). \tag{49}$$

The criterion function in (47) can be minimized using standard optimization algorithms for a range of values for ω and ω_g using penalty function methods for the restrictions $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Initial conditions can be obtained from the "first stage" GMM where $\boldsymbol{\Omega}(\boldsymbol{\theta}) = \mathbf{I}$. This optimization is, typically, easy to implement and its numerical performance in terms of speed has been found quite satisfactory.

Our computational experience suggests that while various small values of ω do not make a large difference, this is not the case for different values of ω_g. Therefore, we follow a cross-validation approach to the selection of this parameter. One firm (with T temporal observations) is omitted at a time and the criterion in (47) is minimized. The observations for the omitted firm are "predicted" at the final parameter values and the value of ω_g is selected so that the sum of squares between the actual and predicted is minimum. We have found that this cross-validation criterion works well and different values of the parameter ω_g result in a rather "deep" U-shaped curve which facilitates the choice of the optimal parameter. With regard to (47), we find that the optimal value for ω_g is 10^3 while the optimal value for ω is 10^{-5}, although results are not sensitive to other values such as 10^{-4}, 10^{-6}, and 10^{-7}.

We also apply cross-validation with respect to another parameter, λ. The covariance matrix is replaced by $\boldsymbol{\Omega}(\boldsymbol{\theta}) + \tilde{\lambda}\mathbf{I}$, since $\boldsymbol{\Omega}(\boldsymbol{\theta})$ turned out to be (numerically) singular in several instances. This happens often in GMM. We set $\tilde{\lambda} = 10^{-7}$. Our instrument set consists of predetermined variables as explained in the data section. Since our model is over-identified, we compute the validity of the over-identifying restrictions.

In some cases local optimization does not work very well suggesting the existence of a multimodal criterion function especially when we choose to work with the GMM-CUE version instead of either (i) one-step GMM only, or (ii) two-step GMM.[6] Therefore, we

[6]One-step GMM estimates parameters based on an initial weight matrix and no updating of the weight matrix is performed except when calculating the appropriate variance-covariance matrix. Two-

use simulated annealing to locate a global optimum of the criterion function.

We have found it useful to experiment also with a bootstrapped version of GMM-CUE to avoid well-known difficulties with the finite-sample distribution of GMM being quite different from its asymptotic normal distribution. This allows us to derive bootstrapped versions of the inefficiency measures outlined above. We employ the pairs block bootstrap.

5 Data

The sample consists of an unbalanced panel, subject to attrition, of at most 77 privately-owned electric utilities (whose names are available upon request from the authors) operating in the U.S. over the period 1988-2005, for a total of 1201 observations. A balanced panel would have yielded 1386 observations. We examine only fossil-fuel-based steam generation. Further, we include a full set of 77 firm-specific dummies and omit the intercept in the quadratic DDF, equation (20).

We model the use of three good inputs (energy, labor, and capital) and one bad input (sulfur) to produce two good outputs (residential and industrial/commercial electricity generation) and three bad outputs (sulfur dioxide (SO_2), carbon dioxide (CO_2), and nitrogen oxide (NO_x)).[7] We also calculate actual prices for the good inputs and good outputs. The price of energy is computed as a weighted average of the cost per million Btu of each fuel, while the price of labor is the wage rate. The price of capital is the yield of the firm's latest issue of long-term debt adjusted for appreciation and depreciation of capital. The prices of residential and industrial/commercial production are derived as total revenues in each category divided by total sales in that category. Data are available

step GMM obtains parameter estimates based on the initial weight matrix, computes a new weight matrix based on those estimates, and then reestimates the parameters based on that weight matrix. Finally, GMM-CUE obtains parameter estimates based on the initial weight matrix, computes a new weight matrix based on those estimates, reestimates the parameters based on that weight matrix, and continues this process to convergence.

[7] Residential electricity is 110 volts as distributed exclusively to residential users, primarily using single-phase wiring. The few residential appliances that require 220 volts (such as water heaters and heat pumps) combine two 110 voltage phases within the home. Industrial/commercial electricity is typically 220 and 110 volts and is sold exclusively using three-phase wiring, which provides a smoother form of electricity that makes large machinery run more efficiently and longer. Thus, the two goods are clearly distinct.

on the quantities, but not prices, for the bad input (sulfur) consumed and bad outputs (CO_2, SO_2 and NO_x) generated by the firm.

In addition to utilizing time dummy variables, we control for firm vintage, τ. This variable and the time dummies are assumed to be separable from the other inputs and outputs.

In rare cases we encountered missing data for some variables. Whenever necessary we accounted for such data by either using the value of the previous period or the average of the previous and the subsequent period, depending on how related variables changed. All continuous variables are standardized to eliminate the sensitivity of results to the scale of the data. For a full discussion of this issue and a more complete description of the data see APT (2016).

All of our empirical results are based on GMM-CUE estimation with a given set of instruments. This set includes time dummies, firm dummies, and all actual (but not latent) prices for good inputs and good outputs, complemented with their squares and their interactions. As indicated above, we assume that utilities are price takers in input markets and are subject to output prices set by regulatory agencies or market forces (if restructured). The vast majority of the utilities in our sample are observed during regulated rather than restructured periods.

6 Empirical Results

Since our model is over-identified, we compute the J-test of the validity of the over-identifying restrictions. This statistic has a p-value of .55 for the full specification of (30) and (31). If we employ only the first terms in these equations, the J-statistic has a p-value of .32, and if we employ only the first two terms, the J-statistic is .47. Thus, firm-specific parameters are estimated using the full specification.

Figs. 2-5 provide summary statistics for the GMM-CUE method. Fig. 2 gives results for price and technical efficiencies of our sample firms. For Case A (with input-specific distortion parameters which are the same for all firms) the median TE is about .84 and for Case B (with firm-specific distortion parameters) median TE is about .85. For Cases A and B, the median AE is about .87 and .82, respectively. Compared to the results in APT (2016), TE is about the same and AE is slightly higher. Differences are not great either between Cases A and B in the current paper or between the results in APT (2016) and those in the current paper.

In Fig. 3 we consider PC, TC, and EC for Cases A and B. For Case A, median posterior values of PC, TC, and EC are very close to zero. The dispersion of PC is greater than that of EC and TC. With Case B, while the median value of TC is approximately zero, those of TC and PC are about .01. Clearly, the Case B results are closer to those in APT (2016) than are the Case A results.

In Fig. 4 we report Case A results for the percent changes in inputs and outputs moving from their current levels with price inefficiency (where all $k_q = 0$) to a state of price efficiency (where all k_q are estimated subject to PM). The results, with capital increasing and energy decreasing are of opposite signs from those in APT (2016). Further, in APT (2016), only SO_2 increases moving to a position of efficient allocation of resources. By contrast, in the present study, SO_2 and NO_x increase, while CO_2 and sulfur decrease.

Fig. 5 presents the same statistics for Case B. The median percent change in energy is about .05, which is positive as with APT (2016), although somewhat smaller. The increases in electricity output for residential and industrial/commercial have median values of approximately .02 and .03, respectively. Median decreases for capital and labor are about -.035 and -.04, respectively. The signs and magnitudes of these changes are nearly identically to those found in APT (2016) and are consistent with Fowlie (2910), who argues that many regulated electric utilities earn super-normal rates of return on capital. Thus the inefficient quantity of capital exceeds its efficient level. Other magnitudes found in the current paper are somewhat smaller than those in APT (2016). We find that the median percent change in NO_x is about -.02, whereas in APT (2016) the this figure is about -.1. In the current paper the median percent change in CO_2 is about -.03, while in APT (2016) this figure is about -1.2. Unlike APT (2016), where the median percent increase in SO_2 is about .08, in the current paper this figure is about -.02. Overall, however, the Case B results are more plausible and closer to those in APT (2016) than are Case A results from Fig. 4. This similarity of results occurs since the estimated distortions in APT (2016) and Case B are both firm-specific.

In Figures 6 and 7 we report bootstrapped finite-sample distributions of distortion parameters for the case of global optimization using two-stage GMM. In Figure 6 the estimator is computed using two-step simulated annealing for global optimization. In Figure 7 we use a local Gauss-Newton procedure (with numerical first- and second-order derivatives and starting values obtained from two-step GMM estimates from the original data set). For both figures, we employ 10,000 replications using the pairs block bootstrap (where the time observations of each firm are kept together), with starting values from the parameter estimates of the two-step GMM in the original sample. We present the empirical distributions for each k where we have averaged across the N bootstrap sampling distributions for each. The results from Fig. 6 for the GMM two-step global estimates provide evidence that the empirical distributions are far from normal, especially the ones that are bimodal. This information is very valuable since standard errors from GMM are well-known to be biased and the finite-sample distribution of GMM is known to be quite different from its asymptotic normal distribution.

This is true (although to a somewhat lesser extent) for the GMM two-step local estimates in Fig. 7. Clearly, in neither case is the assumption of normality justified in the construction of confidence intervals. Otherwise, in terms of mean and dispersion, the results in Fig. 7 are highly similar to those in Fig. 6. In both figures, there is little overlap of the bootstrap distributions with zero, indicating that many significant (although relatively small) distortions of actual quantities from shadow quantities exist in our data. We then compute the minimum p-values among the NT Case B bootstrap estimates of t-values for each distortion parameter to test the null that each coefficient equals zero. These values are reported in Table 1, where the smallest p-value of .00 occurs for energy, capital, residential, industrial, and sulfur. Only the p-values for labor (.13), CO_2 (.17) SO_2 (.22), and NO_x(.23) are greater than .05. Together, these results are consistent with Figs. 6 and 7.

7 Conclusions

Due to its greater flexibility, the DDF has been employed with increasing frequency to estimate multiple-input and multiple-output production, where some inputs and outputs are good while others are bad. Despite this, typically researchers make three restrictive assumptions. First, they a priori assume a direction of movement of firm production toward the frontier in order to measure inefficiency. Second, they assume away the possibility that actual quantities of inputs and outputs are price inefficient. Third, they assume exogeneity of all inputs and all outputs (except for the normalized one).

This paper relaxes each of these assumptions. The first contribution of this paper is that we include parameters to estimate optimal directions which correspond to the firm's PM position. The second contribution is to generalize the DDF to a shadow DDF. This entails adding distortion parameters to each input and output of the DDF, creating shadow quantities. They measure efficient levels of inputs and outputs which correspond to actual prices, subject to PM. Actual levels of inputs and outputs correspond to shadow prices, which we do not measure. To estimate the shadow quantities and the structural parameters, we form the shadow DDF system, which includes the shadow DDF and the first-order price equations from the PM problem. Missing prices for bad inputs and bad outputs are approximated and then used in the first-order conditions for PM. We jointly estimate distortion parameters with the optimal-PM directions which correspond to the firm's PM position.

The third contribution is to estimate the shadow DDF system using a GMM-CUE approach, where we assume that all input and output quantities are potentially endogenous. We use good input and good output prices as part of our instrument set, since their prices are exogenously determined. Our approach is considerably simpler than the Bayesian approach employed in APT (2016).

We compute empirical densities for the parameters and latent variables of our system using an unbalanced panel of 77 U.S. electric utilities for the years 1988-2005. We find modest productivity gains. Bootstrapped finite-sample distributions of distortion parameters indicate relatively small distortions that have little overlap with zero. Using firm-specific estimates of distortion parameters, our results are qualitatively similar to APT (2016), who employ the same data set and compute firm-specific distortions.

References

Atkinson, S.E. and R. Halvorsen, 1976, Interfuel substitution in steam-electric power generation, *Journal of Political Economy* 84, 959-78.

Atkinson, S. E. and D. Primont, 2002, measuring productivity growth, technical efficiency, allocative efficiency, and returns to scale using distance functions *Journal of Econometrics* 108, 203-225.

Atkinson, S.E., D. Primont, and M. Tsionas, 2016, Estimating efficient production with bad inputs and outputs using latent prices and optimal directions, working paper, University of Georgia.

Atkinson, S. E. and M. Tsionas, 2016, Directional distance functions: optimal endogenous directions, *Journal of Econometrics*, 109, 301-314.

Ayres, R.U., Kneese, A.V., 1969. Production, consumption, and externalities. *American Economic Review* 69, 282297.

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

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

Cornwell, C., P. Schmidt, and R.C. Sickles, 1990, Production frontiers with time series variation in efficiency levels, *Journal of Econometrics* 46, 185–200.

Färe, R., S. Grosskopf, and G. Whittaker, 2013, Directional output distance functions: endogenous directions based on exogenous normalization constraints, *Journal of Productivity Analysis* 40,267-269.

Färe, R. and S. Grosskopf 2000, Theory of application of directional distance functions, *Journal of Productivity Analysis* 13, 93-103.

Førsund, F., 2009, Good modelling of bad outputs: pollution and multiple-output production, *International Review of Environmental and Resource Economics* 3, 1-38.

Fowlie, M., 2010, Emissions trading, electricity restructuring, and investment in pollution abatement, *American Economic Review* 100, 837-69.

Hampf, B. and J. Krüger, 2015, Optimal directions for directional distance functions: an exploration of potential reductions of greenhouse gases, *American Journal of Agricultural Economics* 97, 920-938.

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

Kumbhakar, S. 1992, Price distortions and resource-use efficiency in indian agriculture: a restricted profit function approach, *Review of Economics and Statistics*, 74,231-239.

Kumbhakar, S. and E.G. Tsionas 2005, Measuring technical and allocative inefficiency in the translog cost system: a bayesian approach, *Journal of Econometrics*, 126, 355–384.

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

Zofio, J. L., J. T. Pastor, and J. Aparicio, 2013, The directional profit efficiency measure: on why profit inefficiency is either technical or allocative, *Journal of Productivity Analysis* 40, 257-266.

8 Tables and Figures

Fig. 1: Movement from Interior Point to Profit-Maximizing Position

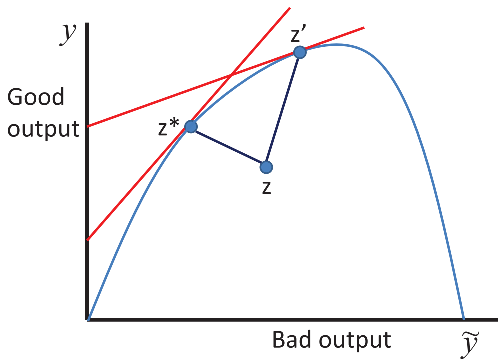

Fig. 2. Sample-mean technical and allocative efficiency

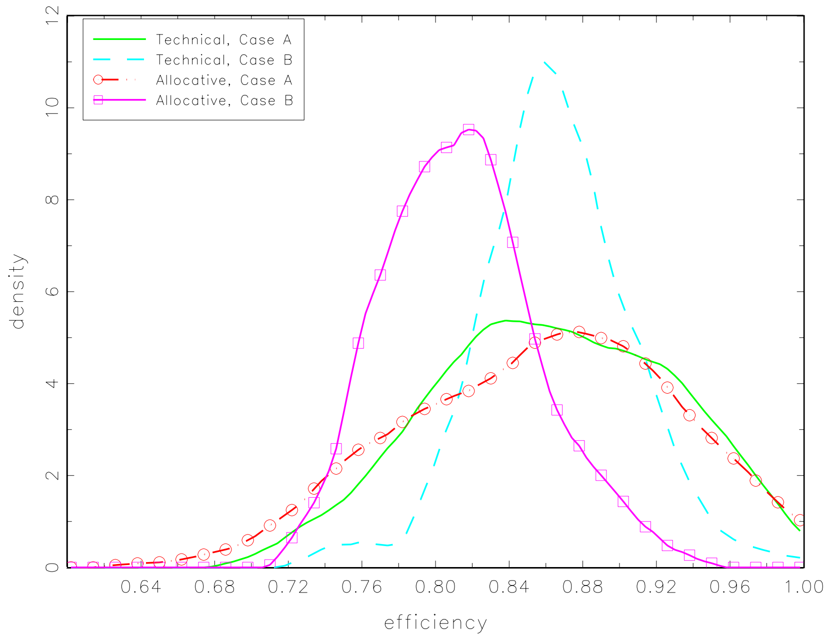

Fig. 3. Sample-mean technical efficiency and productivity change

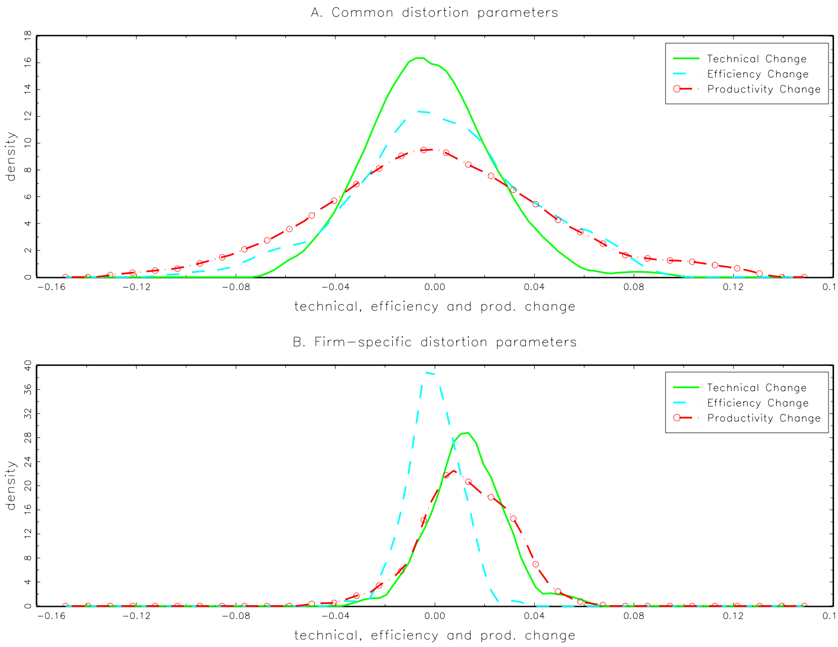

Fig. 4. Input / output changes (sample distributions) for common distortion parameters

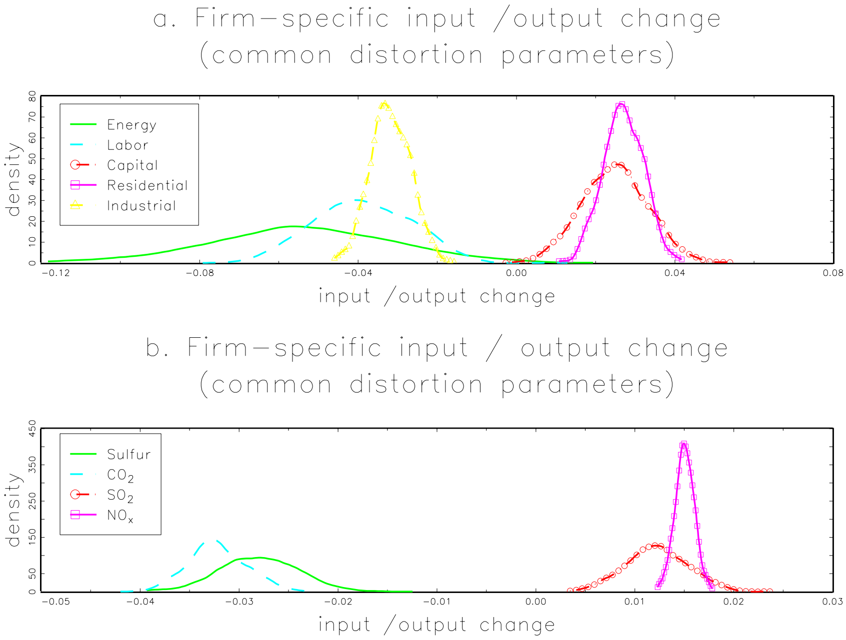

Fig. 5. Input / output changes (sample distributions) for firm-specific distortion parameters

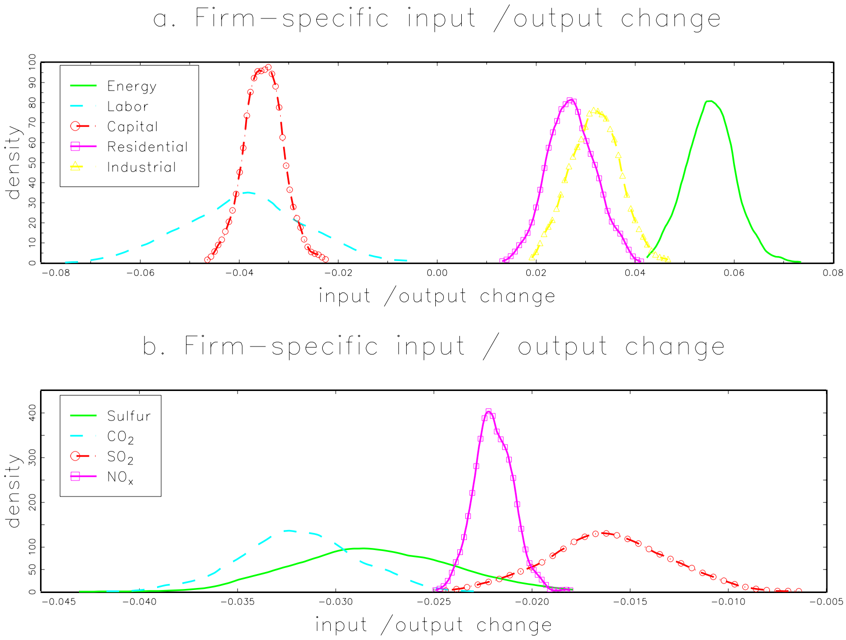

Fig. 6. Bootstrapped distributions, two-stage GMM, global optimization

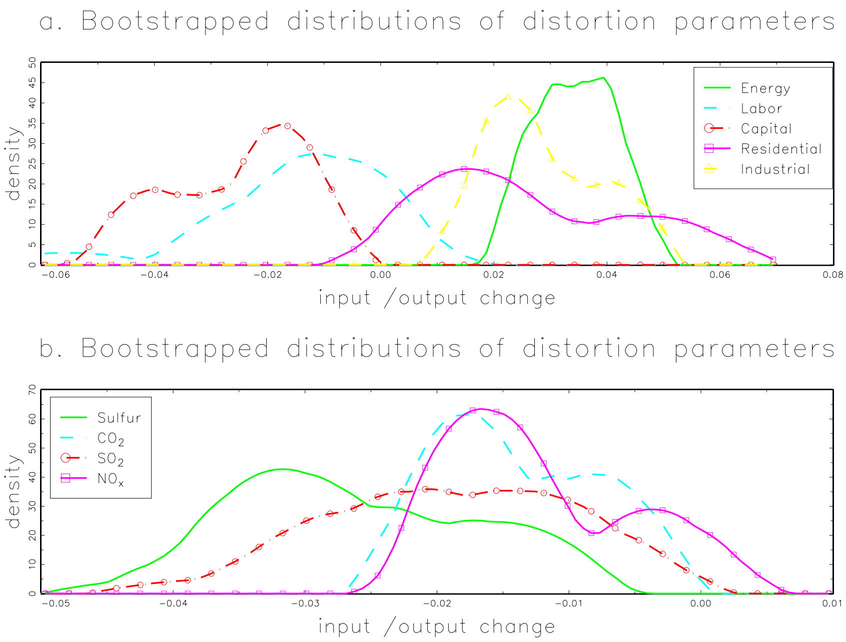

Fig. 7. Bootstrapped distributions, two-stage GMM, local optimization

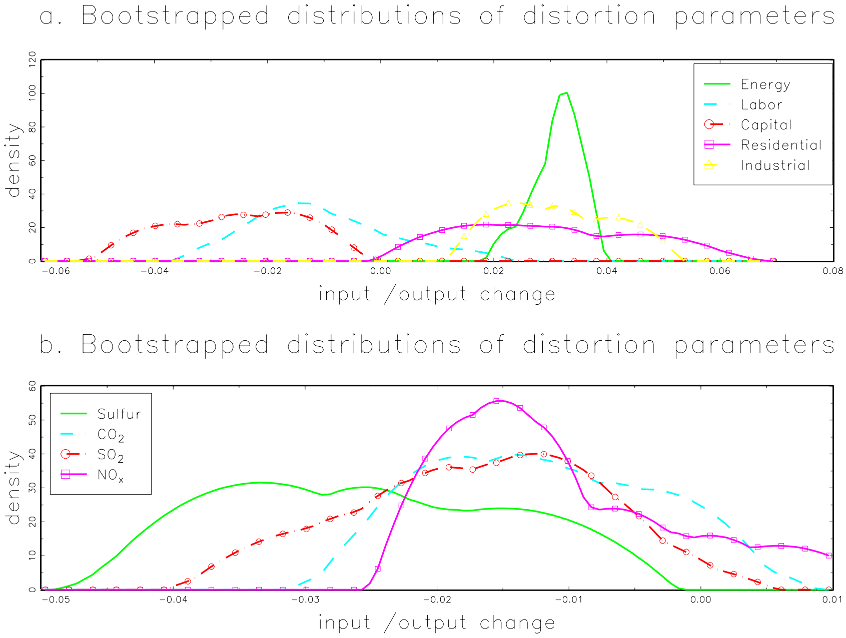

Table 1. Minimum p-values for Bootstrap t-values for Distortion Parameters from Fig. 6

variable	minimum p − value
energy	.00
labor	.13
capital	.00
residential	.00
industrial	.00
sulfur	.00
CO_2	.17
SO_2	.22
NO_x	.23

9 Appendix: Pseudo-code to implement our GMM Algorithm

We follow the standard GMM procedure of specifying a GMM weight matrix initially and then updating this weight matrix on successive iterations. For a given parameter vector $\boldsymbol{\theta}$, choose ω and ω_g, using values suggested above.

1. Solve (20), the directional distance function, in terms of its error term, ϵ_{it}.
2. Solve (24) - (27) in terms of their error terms, $v_{q,it}, q = n, m, j, l$.
3. Insert (30) and (31) into (28) and (29).
4. Insert (28) and (29) into (26) and (27).
5. Specify the matrix of instruments, $\mathbf{Z}$.
6. Compute $\mathbf{G}(\boldsymbol{\theta}; \mathbf{Z})$ in (46) using $\Omega(\boldsymbol{\theta}) = \mathbf{I}$, an identity matrix.
7. Use the previously computed penalty functions using ω and ω_g, and formulate the minimization problem in (47).
8. Solve this problem using a minimization algorithm.
9. After convergence use the $\Omega(\boldsymbol{\theta}) + \lambda\mathbf{I}$ matrix in (48) (for numerical stability, $\lambda = 10^{-7}$) and solve the minimization problem in (47) again.

We have used a standard conjugate-gradients algorithm to solve the optimization problems in steps 8 and 9. This algorithm can be written easily in R, Gauss, Matlab, C, or Fortran.