⊙

Directional distance functions: optimal endogenous directions

Scott E. Atkinson
Department of Economics
Athens, GA 30602,
Tel: 706 549 0977
email: atknsn@uga.edu

Efthymios G. Tsionas
Department of Economics
Athens University of Economics and Business
76 Patission Street, 104 34 Athens, Greece
Tel.: (++30210) 8203338, Fax: (++30210) 8203301, email: tsionas@aueb.gr

Abstract

A substantial literature has dealt with the problem of estimating multiple-input and multiple-output production functions, where inputs and outputs can be good and bad. Numerous studies can be found in the areas of productivity analysis, industrial organization, labor economics, and health economics. While many papers have estimated the more restrictive output- and input-oriented distance functions, here we estimate a more general directional distance function. A seminal paper on directional distance functions by Chambers (1998) as well as papers by Färe, Grosskopf, and Chung (1997), Chambers, Chung, and Färe, (1998), Färe and Grosskopf (2000), Grosskopf (2003), Färe et al. (2005), and Hudgins and Primont (2007) do not address the issue of how to choose an optimal direction set. Typically the direction is arbitrarily selected to be 1 for good outputs and -1 for inputs and bad outputs. By estimating the directional distance function together with the first-order conditions for cost minimization and profit maximization using Bayesian methods, we are able to estimate optimal firm-specific directions for each input and output which are consistent with allocative and technical efficiency. We apply these methods to an electric-utility panel data set, which contains firm-specific prices and quantities of good inputs and outputs as well as the quantities of bad inputs and outputs. Estimated firm-specific directions for each input and output are quite different from those normally assumed in the literature. The computed firm-specific technical efficiency, technical change, and productivity change based on estimated optimal directions are substantially higher than those calculated using fixed directions.

1. Introduction

A large literature has dealt with the problem of estimating multiple-input and multiple-output production functions, where inputs and outputs can be good and bad. Numerous studies can be found in the areas of productivity analysis, industrial organization, labor economics, and health economics. Many studies in the area of child health estimate reduced form equations, thereby avoiding the direct estimation of disaggregated multiple-input, multiple-output structural equations, as summarized in Agee, Atkinson, and Crocker (2012). Another area of extensive study has been firm efficiency, where some researchers assume an aggregate production technology, as in Fernandez, Koop, and Steel (2005).

Other researchers have directly estimated disaggregated multiple-input, multiple-output production functions (structural equations) using *distance* and *directional distance* functions in an attempt to measure the tradeoffs among inputs and outputs, without employing separability or aggregation assumptions. Using an output-oriented *distance function, the* researcher has two options. He can take the approach of holding inputs constant and scale bad outputs and good outputs by the same parameter to reach the production frontier. Pittman (1983) shows that this credits the firm for increasing a bad output (say pollution) along with a good output (say electricity). Alternatively, the researcher can hold constant bad outputs and inputs and measure the distance from the frontier using a proportional upward scaling of good outputs. However, no credit is given for a simultaneous reduction in bad outputs or inputs.

Similarly, using an input-oriented *distance function,* as with Atkinson and Dorfman (2005), a researcher has two options. First, he can hold constant good and bad outputs and scale back all inputs proportionally to reach the frontier. However, again no credit is given for any simultaneous increase (reduction) in good (bad) outputs. Alternatively, one can hold good outputs constant, treat bad outputs like inputs, and scale back both by the same factor of proportionally. However, the equal-proportionality assumption provides no credit for an increase in good outputs. None of these methods credits the firm for simultaneous changes in all inputs and outputs.

The less-restrictive *directional distance* function allows calculation of the distance to the production frontier using different directions for each input and output, both good and bad. If non-zero directions are used to change only inputs (outputs) when measuring productivity

growth, the directional distance function is input- (output-) oriented. When non-zero directions are used for inputs and outputs, the directional distance function is technology-oriented. The choice of the direction of movement of a firm toward the frontier clearly affects measures of all structural parameters as well as the distance from a multiple-output production frontier. This becomes the basis for computing technical efficiency (TE), using an input-, output-, or technology-oriented approach, as well as productivity change (PC), which is the sum of the outward shift of the frontier, termed technical change (TC), and the extent to which the firm catches up to the frontier, termed efficiency change (EC). The latter equals the change in TE.

   The exact direction chosen may substantially affect the calculation of TE as well as the measures comprising PC. As shown by Vardanyan and Noh (2006) and Agee, Atkinson, and Crocker (2012), the parameter estimates depend on the choice of the directional vectors. However, a seminal paper on directional distance functions by Chambers (1998), as well as papers by Färe, Grosskopf, and Chung (1997), Chambers, Chung, and Färe, (1998),   Färe and Grosskopf (2000), Grosskopf (2003), Färe et al. (2005), and Hudgins and Primont (2007) do not address the issue of how to choose an optimal direction set.   As is typical with the empirical applications for each of these studies, Färe et al. (2005) estimated an output-oriented directional distance function for electric utilities involving three good inputs, one good output, and one bad output using directions of +1 for the good output and -1 for the bad output and all inputs.   As a generalization of this approach, Agee, Atkinson, and Crocker (2012) considered the impact of four distinct sets of directions on the estimated parameters of an output-oriented directional distance function, employed to explain measures of child health. While this provides information regarding sensitivity of model results to the directions chosen, the choice amongst them is left to the researcher. No framework has been provided within which to determine an optimal set of directions in a stochastic framework, although in a non-parametric framework Färe, Grosskopf, and Whittaker (2013) make the choice of the direction vector endogenous.

   Feng and Serletis (2014) propose a primal Divisia-type productivity index that incorporates undesirable outputs in a directional distance function with fixed directions of (+1,-1) for good and bad outputs, respectively. However, the paper contains numerous restrictive assumptions, as indicated by Tsionas, Malikov, and Kumbhakar (2014), which substantively affect results once they are generalized. Below we also discuss erroneous

statements about required transformations for estimating distance and directional distance functions. We also note that the Feng and Serletis (2014) proposed aggregation index for productivity growth in their equation (4) is unit-sensitive and therefore is an improper aggregator of good and bad quantities.

In this paper we generalize the previous approaches by estimating the set of directions that is consistent with cost minimization and profit maximization. In order to accomplish this, we first formulate the restrictions that impose the fundamental translation property for input-, output-, and technology-oriented stochastic directional distance functions so that these restrictions contain the directions applied to each input and output. The translation property, akin to the property of linear homogeneity in input (output) quantities for an input (output) distance function, allows one to examine the effect of different directions of movement toward the frontier for different categories of inputs and outputs, both good and bad. We then generalize the dual relationship between the profit function and the technology-oriented directional distance function, as established by Chambers (1998), by assuming profit-maximizing behavior and deriving associated price equations for each input and output. These equations relate their prices to first-order partial derivatives of the directional distance function with respect to the quantity of each input and output and allow identification of directions for each input and output. This set of equations specializes to a system which models cost-minimizing behavior by utilizing only the associated price equations for each input.

We utilize our technique to model the electric utility production process using a set of inputs and outputs, both good and bad. Good inputs are energy (E), labor (L), and capital (K), which includes the annualized capital expenditures on environmental control for the two major restricted air pollutants, sulfur dioxide (SO_2) and nitrogen oxide (NO_X). These capital expenditures are for scrubbers to reduce SO_2 and NO_X emissions and/or modifications of combustion processes to reduce NO_X creation. We also include the sulfur content of fuels, S, as a bad input.[1] This generalization is important, since bad outputs can be reduced by switching to fuels with lower S, as well as modifying combustion processes or installing emission control devices. Trade-offs among these options have not been

---

[1] The study by Yaisawarng and Klein (1994) includes S, SO_2 emissions, electricity generation, and the required good inputs --- production capital, fuel, and labor. However, they exclude the capital cost of pollution control equipment and the emissions of the other two major pollutants.

modeled in any previous study of electric utilities. Further, as bad outputs we include emissions of the three major pollutants --- SO_2, carbon dioxide (CO_2), and NO_X. Since the emissions of CO_2 have never been regulated, typically, studies have included only SO_2 as a bad output. Good outputs are residential (R) and industrial/commercial (IC) electricity generation.

Using a panel of 77 U.S. privately-owned firms producing steam-electric power over 10 years, we jointly estimate a quadratic technology-oriented directional distance function and a set of first-order conditions from the dual cost-minimization and profit-maximization models. The typical fixed-directions approach relies on the assumed directions of (+1,-1) for good outputs and good inputs/bad outputs, respectively. However, we argue that since goods and non-marketed bads are produced by utilities, their relative valuation may not be 1-to-1 for all firms, when we model them as cost minimizers or profit maximizers. Since our data contain input and output price data, we append price equations (where prices are related to marginal products) for inputs to our directional distance function to obtain a cost-minimization directional distance system and the price equations for all good inputs and outputs to obtain a profit-maximization directional distance system. We identify the directions for bad inputs and bad outputs, which lack prices, using methods explained below. Using Markov Chain Monte Carlo (MCMC) methods we estimate these systems, obtaining estimates of all structural parameters, optimal directions, measures of TE, PC, TC, and EC, and estimates of the implied optimal percent changes in inputs and outputs. These directions are those that would prevail in the industry if firms were cost minimizers or profit maximizers. That is, we are estimating directional distance functions, not with directions chosen a priori, but with optimal directions chosen that are consistent with cost minimization or profit maximization.

As we show in our empirical application, the estimated optimal directions imply considerably larger measures of efficiency and productivity change than obtained using the fixed-directions approach. Optimal directions also imply that for the average firm to achieve cost minimization (i.e., be allocatively efficient), it must reduce K relative to L and E. To achieve profit maximization (assuming that pollutant emission levels are given), it must additionally reduce R and IC output.

The remainder of this paper is organized as follows. In section 2, we present the properties of the directional distance function including price equations derived from the cost-minimization and profit-maximization models. We develop the econometrics of directional technology distance functions with fixed directions in section 3. In section 4 we explain our MCMC inference procedure. Data and econometric results follow in section 5 and conclusions in section 6.

2. The Model

2.1 Model Properties

Consider the typically employed a quadratic function of all inputs and outputs as

$$F(z) = \sum_{w=1}^{W} \alpha_w z_w + \tfrac{1}{2} \sum_{w=1}^{W} \sum_{w'=1}^{W} \beta_{ww'} z_w z_{w'} \tag{1}$$

where $\mathbf{z} = [z_1, ..., z_W]'$ includes collectively all inputs and outputs (good or bad). We begin with a simple reformulation of the definition of the directional distance function (see eq. 1 in Hudgins and Primont, 2007 (hereafter HP)):

$\vec{D}(\mathbf{z}, \mathbf{g}) = \max \{\beta : \mathbf{z} + \beta \mathbf{g} = (x + \beta g_x, y + \beta g_y) \in T \}$, with the convention that the directions $g_y > 0, g_x < 0$. The directional distance function has the following properties:

D1: Translation: $\vec{D}(\mathbf{z} + \alpha \mathbf{g}, \mathbf{g}) = \vec{D}(\mathbf{z}, \mathbf{g}) - \alpha$

D2: g-Homogeneity of degree minus one: $\vec{D}(\mathbf{z}, \lambda \mathbf{g}) = \lambda^{-1} \vec{D}(\mathbf{z}, \mathbf{g})$

D3: Good Input Monotonicity: $x' \geq x \rightarrow \vec{D}(x', \tilde{x}, y, \tilde{y}; \mathbf{g}) \geq \vec{D}(x, \tilde{x}, y, \tilde{y}; \mathbf{g})$

D4: Bad Input Monotonicity: $\tilde{x}' \geq \tilde{x} \rightarrow \vec{D}(x, \tilde{x}', y, \tilde{y}; \mathbf{g}) \geq \vec{D}(x, \tilde{x}, y, \tilde{y}; \mathbf{g})$

D5: Good Output Monotonicity: $y' \geq y \rightarrow \vec{D}(x, \tilde{x}, y', \tilde{y}; \mathbf{g}) \leq \vec{D}(x, \tilde{x}, y, \tilde{y}; \mathbf{g})$

D6: Bad Output Monotonicity: $\tilde{y}' \geq \tilde{y} \rightarrow \vec{D}(x, \tilde{x}, y, \tilde{y}'; \mathbf{g}) \geq \vec{D}(x, \tilde{x}, y, \tilde{y}; \mathbf{g})$

D7: Concavity: $\vec{D}(x, \tilde{x}, y, \tilde{y}; \mathbf{g})$ is concave in $(y, \tilde{y}) \in P(x, \tilde{x})$.

D8: Non-negativity: $D(x,\tilde{x},y,\tilde{y};\mathbf{g}) \geq 0$ iff $(y,\tilde{y}) \in P(x,\tilde{x})$.

From HP, the directional distance function in (1) must satisfy D1-D8. In equations (18)-(21), HP derive the restrictions that must be imposed to satisfy D1. Condition D2 is automatically satisfied so long as D1 is imposed and (1) is estimated with zero on the left-hand side. The equations that impose D1 are the only equations that involve the directions (they do not initially appear in equation (1)). To see this suppose for simplicity that

$\mathbf{z} = \left[y_1,...,y_M, x_1,...,x_N \right]'$, where the y's are good outputs, while the x's are good / bad

inputs and bad outputs, so $W = M + N$. Suppose also that $\mathbf{g} = \left[g_1,...,g_M, g_{M+1},...,g_W \right]'$,

where $g_1,...,g_M \geq 0$, and $g_{M+1},...,g_W \leq 0$. The restrictions from HP which satisfy D1 are:

$$\sum_{w=1}^{W} \alpha_w g_w = 1, \tag{2}$$

$$\sum_{w'=1}^{W} \beta_{ww'} g_w = 0, \ \forall w = 1,...,W. \tag{3}$$

Now the mechanics of the model are such that only the restrictions above change when one

selects a particular direction vector $\mathbf{g} \in \mathbb{R}_+^W$.

2.2 Imposing the restrictions

In order to impose the restrictions in (2)-(3) we must rewrite them in terms of some

parameter.

From (2) we have $\alpha_W = \dfrac{1 - \sum_{w=1}^{W-1} \alpha_w g_w}{g_W}$

while from (3) we have

$$\beta_{wW} = -\sum_{w'=1}^{W-1} \beta_{ww'} g_{w'} / g_W, \quad \forall w = 1,...,W$$

so that we place the identically doubly-subscripted parameters on the left-hand side.

Substituting these expressions in (1) we obtain:

$$\vec{D}(\mathbf{z},\mathbf{g}) = \frac{z_W}{g_W} + \sum_{w=1}^{W-1}\alpha_w\left(z_w - \frac{g_w}{g_W}\right) + \tfrac{1}{2}\sum_{w=1}^{W-1}\sum_{w'=1}^{W-1}\beta_{ww'}\left(z_w z_{w'} - \frac{g_w}{g_W}z_w z_W\right)$$

This can also be expressed as:

$$\vec{D}(\mathbf{z},\mathbf{g})g_W = z_W + \sum_{w=1}^{W-1}\alpha_w\left(z_w g_W - g_w\right) + \tfrac{1}{2}\sum_{w=1}^{W-1}\sum_{w'=1}^{W-1}\beta_{ww'}\left(z_w z_{w'}g_W - g_w z_w z_W\right)$$

$$\vec{D}(\mathbf{z},\mathbf{g})g_W = z_W + \sum_{w=1}^{W-1}\alpha_w\left(z_w g_W - g_w\right) + \tfrac{1}{2}\sum_{w=1}^{W-1}\sum_{w'=1}^{W-1}\beta_{ww'}\left(z_w z_{w'}g_W - g_w z_w z_W\right)$$

2.3 First-order conditions

The profit function has been shown to be dual to the directional distance function by Chambers (1998).   He then postulates that the profit-maximizing firm solves

$$\sup\{\mathbf{p}_y(\mathbf{y}+\vec{D}\mathbf{g}_y) - \mathbf{p}_{\tilde{y}}(\tilde{y}+\vec{D}\mathbf{g}_{\tilde{\mathbf{y}}}) - \mathbf{p}_x(\mathbf{x}+\vec{D}\mathbf{g}_x) - \mathbf{p}_{\tilde{x}}(\tilde{x}+\vec{D}\mathbf{g}_{\tilde{x}})\}. \quad (4a)$$

By taking partial derivatives of the previous expression, he derives the first-order conditions for an interior solution to the profit-maximization problem:

$$\frac{p_w}{\mathbf{p}'(J\odot\mathbf{g})} = J\odot\nabla_w\vec{D}(\mathbf{z};\mathbf{g}) = J_w\left(\alpha_w + \sum_{w'=1}^{W}\beta_{ww'}z_{w'}\right), \quad \forall w=1,...,W, \qquad (4b)$$

or equivalently

$$\frac{J_w p_w}{\mathbf{p}'(J\odot\mathbf{g})} = \alpha_w + \sum_{w'=1}^{W}\beta_{ww'}z_{w'}, \quad \forall w=1,...,W,$$

$$(4c)$$

where $J_w=1$ if $w\in\{1,....,M\}$ (that is z_w is a good output), and $J_w=-1$ otherwise, and these conditions are valid only for the elements of $\mathbf{z}$ for which prices are available. Here $\mathbf{p}\in\mathbb{R}^S_{++}$ denotes the price vector.[3] This term is the sum of direction-weighted revenues and cost, and can be interpreted as the Lagrange multiplier.

---

[3] Note that our $\mathbf{p}'(J\odot\mathbf{g})$ is the same direction vector as used by HP except that their directions are defined to be strictly positive. This implies that their version of D1, the

Also $J = \left[J_s, s \in \mathbb{S} \right]'$, where $\mathbb{S}$ is the subset of $\{1,...,W\}$ indicating elements for which we have prices. Let $S = \dim(\mathbb{S})$. So we have prices for only $S < W$ variables, typically only for the good inputs and the good outputs.

One can more accurately write (1) as $\vec{D}(\mathbf{z}; \mathbf{g})$ to emphasize that the estimated distance function is conditional on $\mathbf{g}$. Suppose $\gamma = \begin{bmatrix} \boldsymbol{\alpha} \\ vech(\mathbf{B}) \end{bmatrix}$, where $\boldsymbol{\alpha} = \left[\alpha_1,...,\alpha_W \right]'$, and $\mathbf{B} = \left[\beta_{ww'} \right]_{w,w'=1,...,W}$ denotes all parameters in (1). Since the directions appear only in the restrictions (2) and (3) an even better notation is $\vec{D}(\mathbf{z}; \gamma_{\mathbf{g}})$ to indicate that the parameters γ depend explicitly on the direction $\mathbf{g}$. In matrix notation, we write the restrictions (2) and (3) as:

$$1 + \boldsymbol{\alpha}'\mathbf{g} = 0, \tag{5}$$

$$\mathbf{B}\mathbf{g} = \mathbf{0}_W, \quad \mathbf{B} = \mathbf{B}'. \tag{6}$$

3. The econometrics of directional technology distance functions with fixed directions

Suppose we have fixed a particular vector of directions, say $\mathbf{g} = \left[1, 1, -1, -1, -1 \right]'$. Then the restrictions (5) and (6) provide specific constraints upon the parameters of (1) and (4). Consider the statistical version of (1) and (4) where error terms are added:

$$\vec{D}(\mathbf{z}) = \sum_{w=1}^{W} \alpha_w z_w + \tfrac{1}{2} \sum_{w=1}^{W} \sum_{w'=1}^{W} \beta_{ww'} z_w z_{w'} - v_0 + u, \tag{7}$$

---

translation property, involves positive directions multiplied by the estimated distance for inputs and bad outputs that would be subtracted from the respective initial values, whereas our version of D1 involves negative directions for inputs and bad outputs would be added to respective initial values. The difference between the two formulations of D1 is strictly notational. The denominator of the left-hand-side of (4b) is composed of the sum of positive terms with our formulation. The same is true for their equations (5) and (6).

$$J_w p_w = \nabla_w \vec{D}(\mathbf{z};\mathbf{g})\mathbf{p}'(J \odot \mathbf{g}) = \left(\alpha_w + \sum_{w'=1}^{W} \beta_{ww'} z_{w'} \right) \mathbf{p}'(J \odot \mathbf{g}) - v_w, \ \forall w \in \mathbb{S},$$

(8)

with a one-sided component, u, and standard noise components, v_o and v_w, each with zero mean, reflecting errors in optimization due to random events beyond the control of the firm. This system of equations is a nonlinear simultaneous equation model where the entire vector $\mathbf{z}$ is endogenous, while prices ($\mathbf{p}$) and possibly other variables are predetermined. In total we have W endogenous variables. There are $S < W$ equations in (8). Moreover $\mathbf{p}'\mathbf{g} = \sum_{w \in \mathbb{S}} p_w g_w$.

In our application $S = 5 < W = 9$. At first sight there are S+1 equations above in S endogenous variables. This is not so, because for any direction $\mathbf{g}$ if we multiply the equations of (8) by g_w ($\forall w \in \mathbb{S}$) we get:

$$J_w p_w g_w = \left(\alpha_w g_w + \sum_{w'=1}^{W} \beta_{ww'} z_{w'} g_w \right) \mathbf{p}'(J \odot \mathbf{g}) - v_w g_w.$$

(9)

Summing over $w \in \mathbb{S}$, we have

$$\sum_{w \in \mathbb{S}} v_w g_w = 0.$$

(10)

Therefore, we may drop one equation from (8). Then the distance function in (7) plus the S-1 equations in (8) provide a system of S equations in the S endogenous variables for which we have price data. This leaves another W-S variables without equations but we will deal with this problem by appending a reduced form as in standard treatments of LIML.

Concentrating now on the system (7) and (8) let us write it in somewhat simpler form assuming we have prices for all M good outputs and N good inputs (W=M+N). If necessary, reorder the elements of $\mathbf{z}$ so that the first M elements are good outputs, the next N elements are good inputs, and the remaining elements are bad inputs and bad outputs (for which no prices are available).

$$\vec{D}(\mathbf{z};\boldsymbol{\gamma}_{\mathbf{g}}) = 0 = \sum_{w=1}^{W} \alpha_w z_w + \tfrac{1}{2} \sum_{w=1}^{W} \sum_{w'=1}^{W} \beta_{ww'} z_w z_{w'} - v_0 + u$$

(11)

10

$$J_w p_w = \left(\alpha_w + \sum_{w'=1}^{W} \beta_{ww'} z_{w'} \right) \mathbf{p}' \left(J \odot \mathbf{g} \right) - v_w, \ w = 1,...,S = M + N, \tag{12}$$

where $J_w = 1, \ w = 1,...,M$, $J_w = -1, \ w = M+1,...,M+N-1$ and the translation property restrictions in (5) and (6) have been imposed. Equations (11) and (12) can be written as follows:

$$\boldsymbol{\alpha}'\mathbf{z} + \tfrac{1}{2}\mathbf{z}'\mathbf{B}\mathbf{z} = v_0 - u, \tag{13}$$

$$\mathbf{J}^d \cdot \mathbf{p} = \left(\mathbf{p}' \left(\mathbf{g} \odot J \right) \right) \cdot \left(\boldsymbol{\alpha}_{\mathbf{g}} + \mathbf{B}_{\mathbf{g}} \mathbf{z} \right) - \tilde{\mathbf{v}}. \tag{14}$$

$$\mathbf{z}^* = \mathbf{\Pi}\mathbf{w} + \mathbf{v}^* \tag{15}$$

where, $\mathbf{J}^d = diag\left[J_1,...,J_M, J_{M+1},...,J_{M+N-1} \right]$, and the notation $\boldsymbol{\alpha}_{\mathbf{g}}, \mathbf{B}_{\mathbf{g}}$ emphasizes that the parameters depend on the particular direction. Moreover, we define $\mathbf{p} = \left[p_1,..., p_M, p_{M+1},..., p_{M+N-1} \right]$, and $\tilde{\mathbf{v}} = \left[v_1,..., v_{N+M-1} \right]$. Define also the overall two-sided error term as $\mathbf{v} = \left[v_0, \tilde{\mathbf{v}}' \right]'$. Equation (15) provides the reduced form for $\mathbf{z}^* = \left[z_{M+N+1},..., z_W \right]'$ which denotes all bad inputs and bad outputs, and $\mathbf{w}$ is a $G \times 1$ vector of predetermined variables. In this application we do not actually use any predetermined variables from outside the model, on the grounds that these are, more often than not, difficult to find. Instead we construct $\mathbf{w}$ using exogenous dummy variables as explained in Section 5.1. If $L = W - \left(M + N \right)$ then $\mathbf{z}^*$ is $L \times 1$, and the matrix of reduced form coefficients is $L \times G$. By $\mathbf{v}^*$ we denote the $L \times 1$ vector of errors in the reduced form. Moreover, we assume $\left[v_o, \tilde{\mathbf{v}}', \mathbf{v}^{*'} \right] \sim N_{M+N+L} \left(\mathbf{O}, \boldsymbol{\Sigma} \right)$.

We note that one error in Feng and Serletis (2014), which is not spelled out in Tsionas, Malikov, and Kumbhakar (2014), is their assertion that "Like the radial output distance function, the directional output distance function in (20) [analogous to our equation (11) above] cannot be estimated directly since $\vec{D}_o(z(t),t;g)$ is not observable". This is incorrect, since for the directional distance function, one can obtain exactly the same set of estimated coefficients and standard errors as obtained using the transformation process in Feng and Serletis. This is accomplished by setting the left-hand-side of their equation (20) to

11

zero--which corresponds to efficient firm behavior--imposing the parametric restrictions that guarantee the translation property, and then estimating this restricted version of (20). The advantage of this approach is that the fitted directional distance is directly obtained via estimation of (20) and this estimate can be directly employed in calculations of PC, TC, and EC. The same approach can be taken with the output distance function by setting the left-hand-side of the output distance function (their equation (13)) to 0--since the efficient firm will have $D_o = 1$ --imposing the restrictions for linear homogeneity in outputs, and estimating this restricted equation.   This will yield identical estimated coefficients and standard errors as would be obtained using their proposed normalization.

From the system in (13) - (15) we derive the Jacobian of transformation:

$$\|\mathcal{J}\| = \left\| \begin{matrix} \boldsymbol{\alpha}' + \mathbf{z}'\mathbf{B} \\ \left(\mathbf{p}'\left(\mathbf{g} \odot J\right)\right)\mathbf{B} \end{matrix} \right\| . \tag{16}$$

 If we denote the HP Lagrange multiplier by

$$\lambda_{\mathbf{g}} = \mathbf{p}'\left(\mathbf{g} \odot J\right) > 0,$$

it is clear that (14) becomes:

$$\mathbf{J}^d \cdot \mathbf{p} = \lambda_{\mathbf{g}} \cdot \left(\boldsymbol{\alpha}_{\mathbf{g}} + \mathbf{B}_{\mathbf{g}}\mathbf{z}\right) - \tilde{\mathbf{v}} . \tag{14b}$$

In simplified terms, the equations of this system can be derived from (12) and can be written as:

$$J_w p_w = \left(\alpha_w + \sum_{w'=1}^{W} \beta_{ww'} z_{w'}\right)\lambda_{\mathbf{g}} - v_w , \ w = 1,...,S = M + N \tag{12b}$$

Notice that $J_w p_w$ is $\pm p_w$ depending on whether z_w is input or bad output (so we get $-p_w$) or good output (in which case we get p_w). The right hand side of (12) or (12b) is a linear function of the z_w s, conditional on the HP Lagrange multiplier $\lambda_{\mathbf{g}}$.

Also (16) simplifies to:

$$\|\mathscr{J}\| = \left\| \begin{matrix} \boldsymbol{\alpha}' + \mathbf{z}'\mathbf{B_g} \\ \lambda_g\mathbf{B_g} \end{matrix} \right\| = \left| \alpha_1 + \mathbf{B}'_{g,1}\mathbf{z} \right| \lambda_g^{W-2} \cdot \left\| \mathbf{B}_g^* \right\|, \tag{16b}$$

where α_1 is the first element of $\boldsymbol{\alpha}$, $\mathbf{B}'_{g,1}$ is the first row of $\mathbf{B_g}$, and $\mathbf{B}_g^*$ is the $(W-1)\times(W-1)$ submatrix of $\mathbf{B_g}$ whose first column is deleted. Due to the singularities induced by the constraints we redefine (in connection with (16b) only) $\boldsymbol{\alpha}$ and $\mathbf{z}$ so that their last element is omitted while $\mathbf{B_g}$ is redefined so that its last row is omitted. When $\mathbf{g}$ is arbitrarily selected it is held fixed at specified values. However, these values may not correspond to the cost-minimizing values. When we estimate $\mathbf{g}$ and it is specified to be the same for all firms, there is little or no problem in estimation because of its presence in the Jacobian. However, when we allow $\mathbf{g}$ to differ across firms, standard methods of estimation become problematic. Hence we turn to MCMC based inference.

Of major interest are measures of TE, TC, EC, and PC, computed using the optimal directions relative to fixed directions for all firms. We compute TE_{it} as the average of $\exp(-u_{it})$ over all MCMC draws. In constructing the other measures we follow Chambers (1996) by computing a Luenberger productivity indicator which may be decomposed into EC and TC (see Grosskopf (2003) for additional discussion). Specifically, we compute the percentage change Luenberger productivity change indicator, PC_{it}^L, as

$$\text{PC}_{it}^L = \text{TC}_{it}^L + \text{EC}_{it}^L.$$

Further, we define

$$\text{TC}_{it}^L = .5 \left\{ \frac{\overrightarrow{D}_T^{t+1}(\mathbf{x}_{i,t+1}, \tilde{\mathbf{x}}_{i,t+1}, \mathbf{y}_{i,t+1}, \tilde{\mathbf{y}}_{i,t+1}) - \overrightarrow{D}_T^{t}(\mathbf{x}_{i,t+1}, \tilde{\mathbf{x}}_{i,t+1}, \mathbf{y}_{i,t+1}, \tilde{\mathbf{y}}_{i,t+1})}{A} \right.$$
$$\left. + \frac{\overrightarrow{D}_T^{t+1}(\mathbf{x}_{it}, \tilde{\mathbf{x}}_{it}, \mathbf{y}_{it}, \tilde{\mathbf{y}}_{it}) - \overrightarrow{D}_T^{t}(\mathbf{x}_{it}, \tilde{\mathbf{x}}_{it}, \mathbf{y}_{it}, \tilde{\mathbf{y}}_{it})}{B} \right\}$$

and

$$\text{EC}_{it}^L = \frac{\{\overrightarrow{D}_T^{t}(\mathbf{x}_{it}, \tilde{\mathbf{x}}_{it}, \mathbf{y}_{it}, \widetilde{\mathbf{y}}_{it}) - \overrightarrow{D}_T^{t+1}(\mathbf{x}_{i,t+1}, \tilde{\mathbf{x}}_{i,t+1}, \mathbf{y}_{i,t+1}, \widetilde{\mathbf{y}}_{i,t+1})\}}{C},$$

which are identical to the definitions of Chambers (1996) except that we convert them into percent changes by dividing by A, B, or C, which are the midpoints of the two associated

4. MCMC – based inference

Suppose $\boldsymbol{\theta} \in \mathbb{R}^d$ is the parameter vector, augmented to include parameters for each direction and $p(\boldsymbol{\theta} \,|\, \mathbb{Y}) \propto \mathcal{L}(\boldsymbol{\theta}; \mathbb{Y}) p(\boldsymbol{\theta})$ is the kernel of the posterior distribution where $\mathbb{Y}$ denotes the data. The likelihood function is $\mathcal{L}(\boldsymbol{\theta}; \mathbb{Y})$ and $p(\boldsymbol{\theta})$ denotes the prior. Due to the high nonlinearity of the posterior and the fact that the parameters and the directions appear in the Jacobian, we use the method of Girolami and Calderhead (2011) to perform Bayesian inference. As the authors write "[t]*he proposed methodology exploits the Riemannian geometry of the parameter space of statistical models and thus automatically adapts to the local structure when simulating paths across this manifold providing highly efficient convergence and exploration of the target density*".

Given a kernel posterior distribution the method of Girolami and Calderhead (2011) aims at obtaining draws by exploiting the gradient and the Hessian. Roughly speaking parameters are updated so that they visit the area of high posterior probability mass, not unlike standard Newton-Raphson numerical optimization techniques.

Following Girolami and Calderhead (2011) we utilize Metropolis adjusted Langevin and Hamiltonian Monte Carlo sampling methods defined on the Riemann manifold, since we are sampling from target densities with high dimensions that exhibit strong degrees of correlation. Consider the Langevin diffusion

$$d\boldsymbol{\theta}(t) = \tfrac{1}{2}\nabla \log p(\boldsymbol{\theta}(t); \mathbb{Y}) dt + d\mathbf{B}(t),$$

where $\mathbf{B}$ denotes the d-dimensional Brownian motion. The first-order Euler discretization provides the following candidate generation mechanism:

$$\boldsymbol{\theta}^* = \boldsymbol{\theta}^o + \tfrac{1}{2}\varepsilon^2 \nabla \log p\left(\boldsymbol{\theta}^o ; \mathbb{Y}\right) + \varepsilon \mathbf{z},$$

where $\mathbf{z} \sim N_d\left(\mathbf{0}, \mathbf{I}\right)$, and $\varepsilon > 0$ is the integration step size. Since the discretization induces an unavoidable error in approximation of the posterior, a **Metropolis step** is used, where the proposal density is

$$q\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^o\right) = N_d\left(\boldsymbol{\theta}^o + \tfrac{1}{2}\varepsilon^2 \nabla \log p\left(\boldsymbol{\theta}^o ; \mathbb{Y}\right), \varepsilon^2 \mathbf{I}\right),$$

with acceptance probability $a\left(\boldsymbol{\theta}^o, \boldsymbol{\theta}^*\right) \min\left\{1, \dfrac{p\left(\boldsymbol{\theta}^* \mid \mathbb{Y}\right) q\left(\boldsymbol{\theta}^o \mid \boldsymbol{\theta}^*\right)}{p\left(\boldsymbol{\theta}^o \mid \mathbb{Y}\right) q\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^o\right)}\right\}.$

The Brownian motion of the Riemann manifold is given by:

$$d\tilde{\mathbf{B}}_i\left(t\right) = \left|\mathbf{G}\left(\boldsymbol{\theta}(t)\right)\right|^{-1/2} \sum_{j=1}^{d} \frac{\partial}{\partial \boldsymbol{\theta}_j}\left[\mathbf{G}^{-1}\left(\boldsymbol{\theta}(t)\right)_{ij} \left|\mathbf{G}\left(\boldsymbol{\theta}(t)\right)\right|^{1/2}\right] dt + \left[\sqrt{\mathbf{G}^{-1}\left(\boldsymbol{\theta}(t)\right)} d\mathbf{B}(t)\right]_i,$$

for $i = 1, \ldots, d$.

The discrete form of the above stochastic differential equations is:

$$\boldsymbol{\theta}_i^* = \boldsymbol{\theta}_i^o + \tfrac{1}{2}\varepsilon^2\left[\mathbf{G}^{-1}\left(\boldsymbol{\theta}^o\right) \nabla \log p\left(\boldsymbol{\theta}^o ; \mathbb{Y}\right)\right]_i - \varepsilon^2 \sum_{j=1}^{d}\left[\mathbf{G}^{-1}\left(\boldsymbol{\theta}^o\right) \frac{\partial \mathbf{G}\left(\boldsymbol{\theta}^o\right)}{\partial \boldsymbol{\theta}_j} \mathbf{G}^{-1}\left(\boldsymbol{\theta}^o\right)\right]_{ij} +$$

$$\varepsilon^2 \sum_{j=1}^{d} \mathbf{G}^{-1}\left(\boldsymbol{\theta}^o\right)_{ij} tr\left[\mathbf{G}^{-1}\left(\boldsymbol{\theta}^o\right) \frac{\partial \mathbf{G}\left(\boldsymbol{\theta}^o\right)}{\partial \boldsymbol{\theta}_j}\right] + \varepsilon\left[\sqrt{\mathbf{G}^{-1}\left(\boldsymbol{\theta}(t)\right)}\mathbf{z}\right]_i \triangleq$$

$$\boldsymbol{\mu}\left(\boldsymbol{\theta}^o, \varepsilon\right)_i + \varepsilon\left[\sqrt{\mathbf{G}^{-1}\left(\boldsymbol{\theta}(t)\right)}\mathbf{z}\right]_i.$$

The *proposal density* is $\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^o \sim N_d\left(\boldsymbol{\mu}\left(\boldsymbol{\theta}^o, \varepsilon\right), \varepsilon^2 \mathbf{G}^{-1}\left(\boldsymbol{\theta}^o\right)\right)$ and the acceptance probability

has the standard *Metropolis form*: $a\left(\boldsymbol{\theta}^o, \boldsymbol{\theta}^*\right) = \min\left\{1, \dfrac{p\left(\boldsymbol{\theta}^* \mid \mathbb{Y}\right) q\left(\boldsymbol{\theta}^o \mid \boldsymbol{\theta}^*\right)}{p\left(\boldsymbol{\theta}^o \mid \mathbb{Y}\right) q\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^o\right)}\right\}.$

In our case the parameter vector includes the structural parameters of the distance function as well as the directions. Technical inefficiency is drawn from its standard conditional posterior distribution resulting from (7). It is not possible to factor the

conditional posterior distributions of directions and obtain them in known forms or, at least, forms that are amenable to posterior simulation so we rely on the general Girolami and Calderhead (2011) update.

In particular, we use the following blocking scheme:

a) Draw **u**, a vector all observations of u from (11), conditional on data, **g, Σ** and **θ,**

b) Draw directions **g**, conditional on data, **u, Σ** and **θ**.

c) Draw the remaining parameters, **θ**, conditional on data, **u, Σ** and **g**.

d) Draw **Σ** conditional on data, **u, θ** and **g**.

Steps (b) and (c) are implemented using two steps of the Girolami-Calderhead Riemannian manifold MCMC scheme. In step (a) the conditional distribution is $u \mid \mathbf{g}, \mathbf{\theta}, \mathbf{\Sigma}, \mathbb{Y} \sim N_{+}\left(-\mathbf{a}'\mathbf{z} - \tfrac{1}{2}\mathbf{z}'\mathbf{Bz}, \ \sigma_{v_o}^2\right)$, where $\sigma_{v_o}^2$ denotes the (+1,1) element of $\mathbf{\Sigma}$ whose remaining elements in the first row and column are assumed zero, for simplicity. Therefore, $\mathbf{\Sigma} = \begin{bmatrix} \sigma_{v_o}^2 & \mathbf{0}' \\ \mathbf{0} & \mathbf{\Sigma}_* \end{bmatrix}$. The elements of $\mathbf{\Sigma}_*$ in step (d) can be drawn from an inverted Wishart distribution.

We define our priors as the following. We assume a conditionally conjugate prior inverted Wishart distribution, $iW\left(\bar{v}, \bar{\mathbf{\Sigma}}_*\right)$, for the different elements of $\mathbf{\Sigma}_*$, where we set $\bar{v}$ to 10 and $\bar{\mathbf{\Sigma}}_*$ to $10^{-2}\mathbf{I}_{M+N+L-1}$. For $\sigma_{v_o}^2$ we assume $\dfrac{\bar{q}}{\sigma_{v_o}^2} \sim \chi^2\left(\bar{v}_o\right)$ with $\bar{v}_o = 1$ and $\bar{q} = 0.01$. For the elements of $\mathbf{\theta}$ we adopt a semi-informative prior implementing the theoretical restrictions in stochastic form. The theoretical restrictions amount to $\mathbf{A}\mathbf{\theta} = \mathbf{0}$. The restrictions are imposed using $\mathbf{A}\mathbf{\theta} = \mathbf{\xi}$, where $\mathbf{\xi} \sim N_R\left(\mathbf{0}, 10^{-6}\mathbf{I}\right)$, where R is the number of restrictions. The directions, **g**, are assumed time-invariant and firm-specific. ==In principle, the optimal **g** varies with input/output prices/quantities and therefore is not time-invariant. However, a model with time-varying **g** was too difficult to estimate.== This makes the imposition of the restrictions in stochastic rather than determin[💬]form, even more important: exact imposition of the restrictions results in a system which is highly nonlinear in **θ** and makes programming somewhat cumbersome.

The use of the prior allows estimating directions **g** in an absolute rather than relative manner. The latter would necessitate imposing restrictions of the form $\mathbf{g}'\mathbf{g} = 1$. This theoretical advantage is hampered by the possibility that estimates of **g** can be sensitive to $\underline{\mathbf{V}}$ if the data are not very informative about directions. As we will see this is not the case in our application when we vary $\underline{\mathbf{V}}$ in its plausible range. The benchmark prior is $\mathbf{g} \sim t_\nu\left(\underline{\boldsymbol{\mu}}, 2\mathbf{I}\right)$, where the elements of $\underline{\boldsymbol{\mu}}$ are 1 or -1 for good outputs and bad outputs/inputs, respectively, where t_ν denotes the Student-t distribution with $\nu = 5$ degrees of freedom parameter. The choice of the Student-t is motivated by the need of the weighted bootstrap that the tails of the benchmark prior must be fatter relative to the tails of the priors for which we need to perform sensitivity analysis.

5. Data and Empirical Results

5.1 Data

The sample consists of 77 privately owned electric utilities, listed in Table 1, which operated in the U.S. over the 10-year period 1988-97, for a total of 770 observations. Extending the sample further results in a loss of many firms due to merger and the cessation of operation of fossil fuel-based steam generation. Since technologies for nuclear, hydroelectric, and internal combustion differ from that of fossil fuel-based steam generation and because steam generation dominates total production by investor-owned utilities during the time period under investigation, we limit our analysis to this component.

For our sample of utilities, total kWh output has remained relatively constant over time, SO_2 emissions have fallen dramatically due to a cap-and-trade system, NO_X emissions have declined somewhat, and CO_2 emissions, which are unregulated, have grown over time. Further, the input of S and production K have dropped slightly, the usage of L has dropped considerably, environmental K has increased substantially, and the usage of E has increased marginally. [4]

We construct good inputs and good outputs as follows. First, the quantity of E is the total Btus of fuel consumed and we construct the quantity of L by taking the expenditures on labor, multiplying by the ratio of steam to total net generation (to obtain the portion

---

[4] Details on aggregate inputs and outputs are available from the authors.

attributable to fossil-fuel generation), and dividing by the price of L (defined below). Both variables are obtained from *FERC Form 1*. This form also reports the total of the dollars of interest plus depreciation paid by each utility for production K and pollution control K. We multiply this sum by the ratio of steam to total net generation to obtain our measure of K used in steam generation. We decompose the good outputs, R and IC, from total steam output in mWhs by multiplying the percent of sales revenue in each category times total steam output. These data are again taken from *FERC Form 1*.

We construct the prices of good inputs and outputs in the following manner. The price of E is computed as a weighted-average of the cost per million Btus of each fuel, taken from the Department of Energy, Energy Information Administration (EIA) Form 767 Boiler Files. The price of L is the wage rate, defined as the sum of salaries and wages charged to electric operation and maintenance, divided by the number of full-time plus one half the number of part-time employees, obtained from *FERC Form 1*. The price of K is the product of the yield of the firm's latest issue of long-term debt, adjusted for appreciation and depreciation of capital using the Christensen-Jorgenson (1970) cost of capital formula, and the Handy-Whitman index for electric utility construction costs. These data were taken from *FERC Form 1* and *Moody's Public Utility Manual*. Prices of R and IC are derived as total revenues in each category divided by total sales in that category, where the data were taken from *FERC Form 1.*

The bad input, S, and the bad outputs --- emissions of SO_2, CO_2, and NO_X --- are obtained from the Department of Energy Information Administration (EIA) *Form 767 Boiler Files*. All pollutants are measured in tons.

Since we have panel data, we add dummy variables for time periods and firms to equation (13). These appear separably from input and output quantities to maintain tractability of our model. We also control for firm vintage, which is computed for each firm as the weighted-average age in years, where the weights are the firm's kWh output, taken from *FERC Form 1*. This variable is not interacted with any other inputs or outputs. We define **w** as the set of firm dummies, time dummies, and their interactions, the prices of good inputs and outputs, their squares, and their interactions. In rare cases we encountered missing data for some variables. Whenever necessary we accounted for

missing data points by either using the value of the previous period or the average of the previous and the subsequent period.

5.2 MCMC Modeling

Potentially all variables are endogenous, since they are quantities of inputs or outputs that are under the control of the firm. Referring to eq. (14), we have prices and hence price equations only for good inputs and good outputs. These equations identify these endogenous variables. For the other endogenous bad input and bad outputs we need the reduced form equations in (15) to achieve identification In our MCMC computations, we specify that each structural parameter has prior mean zero with a diffuse prior covariance matrix. In our application the Hessian of the log-likelihood is not available in closed form, so $\mathbf{G}(\mathbf{\theta})$ and $\mathbf{G}^{-1}(\mathbf{\theta})$ are not available. Instead we consider the observed Fisher information matrix $\mathbf{G}(\mathbf{\theta}) = \widehat{\operatorname{cov}}\left[\dfrac{\partial}{\partial \mathbf{\theta}} \log \mathscr{L}(\mathbf{\theta}; \mathbb{Y})\right]$, as in BHHH. The score of the log-likelihood is computed numerically. In all applications we take 600,000 passes, the first 100,000 of which are discarded to mitigate the impact of start-up effects, and then we keep every 10th draw resulting in a total of 50,000 draws which, as we have found, are practically uncorrelated. Inferences for the parameters and functions of interest are based on these 50,000 draws. During the burn in phase, we choose the step size parameter, ε, so that the acceptance rate later on is approximately 80% (see Girolami and Calderhead, 2011, p. 22) resulting in values close to 0.01.

Each of these is a function of both the data (firm- and time- dependent) and the parameters. The functions are recorded for each firm and time period for a given parameter draw of the Riemann Manifold--Metropolis Adjusted Langevin Algorithm (RM-MALA). Suppose a function of interest $\varphi_{it}(D, \theta)$ depends on the data and the parameters θ which include latent variables like the optimal directions. If $\left\{\theta^{(s)}, s = 1, ..., S\right\}$ denotes the available draws, the posterior mean can be computed as $E_{\theta|D}\left[\varphi_{it}(D, \theta)\right] = S^{-1} \sum_{s=1}^{S} \varphi_{it}\left(D, \theta^{(s)}\right)$. Functions of interest may include technical inefficiency and, of course, the optimal directions themselves. Sample distributions of functions of interest are reported using their posterior means and standard deviations so that parameter uncertainty is fully taken into account.

5.3 Posterior Densities

In Table 2 we present the optimal directions computed for five models.   With model I, we calculate optimal directions for the three good inputs, the bad input, and the three bad outputs. In model II we do this for only the good inputs and in Model III for these inputs and CO_2 emissions. Model IV computes optimal directions for all good inputs and good outputs, while Model V computes optimal directions for all inputs and outputs, both good and bad. Thus Model II is our pure cost-minimization model while Models IV is our pure profit-maximization model, assuming that levels of the pollutants and the bad input are held constant.

The posterior mean estimated directions for the good outputs and the set of bad outputs, bad input, and good inputs are quite different from the priors of +1 and -1, respectively, which typically have been imposed in previous studies estimating directional distance functions. Across all five models, the estimated direction for K is negative, while the estimated directions for L and E are positive. This implies that cost-minimizing and profit-maximizing firms (which would no longer be regulated) wish to decrease their usage of K and increase their usage of L and E. The current over capitalization is due to the operation of nearly all of our sample utilities in rate-of-return-regulated jurisdictions, where the regulated rate of return exceeds the market rate. This phenomenon is known as the Averch-Johnson effect, postulated by Averch and Johnson (1962).   With Model IV, the utilities on average will also choose to reduce their production of R and IC since they were forced to produce more good output than is profit-maximizing by the regulatory commissions. Models I and III are cost-minimizing cases with the inclusion of additional equations for S, SO_2, CO_2, and NO_X in Model I and for CO_2 in Model III. Model V examines profit maximization with the additional inclusion of equations for the bad input and all the bad outputs. With Models I, III, and V, since we have no prices for bad outputs and the bad input, their directions cannot be optimized as part of a cost-minimization or profit-maximization solution.   Rather, their directions reflect historical trends exhibited by the reduced-form equations for the bad outputs and the bad input. Thus, with Model V, the historical decreases in S, SO_2, and NO_X determine negative directions. With all models, CO_2 is optimally increased, since it is an unregulated pollutant, which determines a positive direction. With regard to S, in Model I the negative direction for the bad input is similar to that with Model V--firms historically have reduced the use of sulfur due to environmental

20

regulations. Although the directions for SO_2 and NO_X differ in sign between Models I and V, we place more confidence in the latter, since the posterior densities for these two pollutants are considerably more separated from zero than are those of Model I. In fact, Models II-V are more precisely identified in terms of posterior densities and tell a very consistent story of expected changes in inputs and outputs under a change to either cost-minimization or profit-maximization from the current system of rate-of-return regulation.

In Figures 1-5 we illustrate the value of estimating firm-specific optimal directions, focusing on firms 1, 22, 28, 46, and 54 (see Table 1), which are five geographically disperse, large utilities. One burns primarily natural gas, one primarily coal, and the others a mix of fuels. In Figure 1, we present these firms' Model I estimated posterior densities for optimal directions of K, L, and E. Firm-specific information is useful for all of these inputs, both in terms of the relative inter-firm sizes of directions for each input as well as their relative dispersion. In Figures 2-5, we see that firm-specific posterior densities for optimal directions with Models II-V are often well-separated for all good inputs. In addition, for some firms, noticeable differences exist in the location of the posteriors, when compared across the models.

We also present aggregate (over all firms) posterior densities for optimal directions of Models IV and V in Figures 6a-6b. In Figure 6a we show the aggregate estimated posterior densities for optimal directions of K,L,E, R, and IC for Models IV and V, while in Figure 6b we present the additional Model V aggregate posterior densities for the optimal directions of S, SO_2, CO_2, and NO_X. Considerable separation of individual densities is clearly exhibited in both figures.

In Table 3 we examine the implications of cost-minimizing or profit-maximizing behavior in terms of percent changes in inputs and outputs for all models relative to actual quantities. These range from approximately -.07 (in the case of SO_2 in Model V) to approximately .07 (for E in Model IV). Such estimated changes are reasonable in magnitude for a switch from a rate-of-return regime to either profit maximization or cost minimization. In addition, the posteriors of all changes are well-separated from zero. We have already discussed the reasonableness of the signs attached to the estimated changes, which are identical to those in Table 2. With Model IV, the approximate average percent change in K is -.05, in L is .03, in E is .07, in R is -.06, and in IC is -.03. With Model V, the approximate average percent change in K is -.04, in L is .06, in E is .04, in R is -.05, and in IC is .06. The approximate average percent change for S is -.04, for SO_2 is -.06, for CO_2 is .04, and for NO_X is -.06.

21

Figure 7a displays the posterior densities of the percent changes in good inputs and good outputs for Models IV and V, while Figure 7b presents these densities for the bad outputs and the bad input of Model V. With Model IV, the percent changes for K and L are more precisely estimated than those for E, R, and IC. With Model V, R has a more disperse posterior than the other inputs and outputs. In Figure 7b, there is considerable separation of all posteriors from zero, although there is little separation among the posteriors for S, SO_2, and NO_X.

For Models IV and V in Figure 8, we graph TE computed using optimal and fixed directions of (+1,-1) for good outputs and inputs/bad outputs, respectively. Using the optimally estimated directions, the mode of TE is approximately .10 to .15 points higher for these models. From Figure 9, we see that by using optimal directions, TC increases from an average of about -.01 with fixed directions to an average of about .02 for Model IV and from an average of about -.018 with fixed directions to an average of about .025 for Model V. Finally, in Figure 10, we examine the changes in PC due to the use of optimal directions. For Model IV, PC increases from an average just above zero to an average of about .02 and for Model V from an average of -.03 to an average of about .025. The increased levels of TE, TC, and PC are quite dramatic, yet still very reasonable, and cast the electric utility industry in a considerably more efficient and progressive light when viewed using optimal directions.

In Table 4 we present the MCMC convergence diagnostics which indicate that the MCMC process did converge for Models I-V. Specifically we have used Geweke's (1992) convergence diagnostics (GCD) and the relative numerical efficiency (RNE). The GCD convergence diagnostic tests whether the draws for each parameter have converged by computing a t-statistic between the first 20% and last 50% of draws. The variance used in the t-statistic is computed using the value of the spectral density at the origin with an AR(10) approximation. For i.i.d. sampling from the posterior, RNEs should be close to one. Since, unavoidably, MCMC introduces autocorrelation (as indicated by the autocorrelation function (acf)), RNEs should not be very close to zero. As seen from the table, the RNEs are large enough , so we expect that the posterior is adequately explored--although not as efficiently as by an iid sampler, which, unfortunately, we do not have at our disposal. For the RNE and the GCD, we compute 99% intervals by ranking the values of these statistics across all parameters and then constructing 99% intervals around the median. The Girolami-Calderhead sampler seems to overcome the problems that other algorithms (like straightforward implementations of Metropolis-Hastings) would have. This is additional

evidence that the Girolami-Calderhead sampler has great potential in applied econometrics and performs well in the context of MCMC with data augmentation: convergence is achieved in reasonably few iterations, autocorrelation is mitigated, and RNEs are improved drastically.

5.4 Sensitivity analysis

Suppose we have a model whose likelihood function is $L(\boldsymbol{\theta}; \mathbf{Y})$ where $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p$ represents the parameter vector and $\mathbf{Y}$ is the data. Suppose also we have a benchmark prior of the form $p_o(\boldsymbol{\theta})$ --it is not necessary that all integrating constants are present in the likelihood and the prior. The posterior distribution is

$$p_o(\boldsymbol{\theta} \mid \mathbf{Y}) \propto L(\boldsymbol{\theta}; \mathbf{Y}) \cdot p_o(\boldsymbol{\theta}).$$

From the posterior we have obtained an MCMC sample $\left\{\boldsymbol{\theta}^{(s)}, s = 1, ..., S\right\}$. Suppose now we have a different prior $p(\boldsymbol{\theta})$ and the posterior is:

$$p(\boldsymbol{\theta} \mid \mathbf{Y}) \propto L(\boldsymbol{\theta}; \mathbf{Y}) \cdot p(\boldsymbol{\theta}).$$

We are interested in an examination of how the posterior mean $\overline{\boldsymbol{\theta}} = \int_{\Theta} \boldsymbol{\theta} p(\boldsymbol{\theta} \mid \mathbf{Y}) d\boldsymbol{\theta}$ changes as we change the prior $p(\boldsymbol{\theta})$. In more precise notation the prior is denoted by $p(\boldsymbol{\theta}) = p(\boldsymbol{\theta}; \alpha)$ and $p_o(\boldsymbol{\theta}) = p(\boldsymbol{\theta}; \alpha_o)$ so that the priors differ with respect to the parameters α. Here we use the weighted bootstrap method presented by Smith and Gelfand (1992), which is similar to the sampling-importance-resampling method of Rubin (1988). Define

$$\tilde{w}^{(s)} = \frac{p\left(\boldsymbol{\theta}^{(s)}\right)}{p_o\left(\boldsymbol{\theta}^{(s)}\right)} \text{ and } w^{(s)} = \frac{\tilde{w}^{(s)}}{\sum_{s=1}^{S} \tilde{w}^{(s)}}.$$ Next we draw $\boldsymbol{\theta}_*^{(s)}$ from the discrete distribution

whose support is $\left\{\boldsymbol{\theta}^{(s)}, s = 1, ..., S\right\}$ with probabilities $w^{(s)}$. In turn this is, approximately, a sample from $p(\boldsymbol{\theta} \mid \mathbf{Y})$ and the posterior mean is, approximately, $\overline{\boldsymbol{\theta}} = S^{-1} \sum_{s=1}^{S} \boldsymbol{\theta}_*^{(s)}$. The

approximation improves as $S \to \infty$. Care should be taken that the tails of $p_o(\boldsymbol{\theta})$ are fatter relative to any of $p(\boldsymbol{\theta})$s that we are going to examine.

To implement sensitivity analysis we vary the critical parameters of the prior for directions $\mathbf{g} \sim N(\underline{\boldsymbol{\mu}}, \underline{\mathbf{V}})$. Since we use standardized data we vary the elements of $\underline{\boldsymbol{\mu}}$ between -2 and 2 and the diagonal elements of $\underline{\mathbf{V}}$ between 0.1^2 and 3^2. The weighted bootstrap approach is implemented using 1,000 different priors which produce 1,000 different posterior means for the directions $\mathbf{g}$. The disributions of posterior means relative to the posterior mean from the benchmark prior (percentage changes) are reported in Figure 12. The posterior estimates of directions show little sensitivity to the choice of priors.

6. Conclusions

A considerable literature has dealt with the problem of estimating multiple-input and multiple-output production functions, where inputs and outputs can be a mix of good and bad, in the areas of productivity analysis, industrial organization, labor economics, and health economics, among others.   Typically researchers have estimated output- and input-oriented distance functions. However, in this paper we have estimated a more general directional distance function. Although the chosen distances can substantially affect measures of TE, PC, TC, and EC, no published work has addressed the issue of how to choose an optimal direction set. By estimating the directional distance function together with the first-order conditions for cost minimization and profit maximization using Bayesian methods, we estimate optimal directions, which are consistent with cost minimization and profit maximization, respectively, by the firm. We find that these estimated directions vary noticeably from the typically used (+1,-1) for good outputs and inputs/bad outputs. Models with estimated optimal directions generate substantial increases in computed technical efficiency and productivity change relative to the fixed-directions case. Further, the estimated percent changes in input usage and output generated by firms to achieve cost minimization or profit maximization are very reasonable.

Table 1: List of firms

1 Alabama_Power_Co.	40 KGE,_A_Western_Resources_Company
2 AmerenCIPS	41 Long_Island_Lighting_Co.
3 AmerenUE	42 Louisville_Gas_&_Electric_Co.
4 Appalachian_Power_Co.	43 Minnesota_Power_&_Light_Co.
5 Arizona_Public_Service_Co.	44 Mississippi_Power_Co.
6 Atlantic_City_Electric_Co.	45 Montana_Dakota_Utilities_Co.
7 Baltimore_Gas_&_Electric_Co.	46 Montana_Power_Co.
8 Boston_Edison_Co.	47 New_York_State_Electric_&_Gas_Corp.
9 Carolina_Power_&_Light_Co.	48 Niagara_Mohawk_Power_Corp.
10 Central_Hudson_Gas_&_Electric_Corp.	49 Northern_Indiana_Public_Service_Co.
11 CLECO_Corp.	50 Northern_States_Power_Co.
12 Central_Maine_Power_Co.	51 Ohio_Edison_Co.
13 Central_Power_&_Light_Co.	52 Ohio_Power_Co.
14 Cincinnati_Gas_&_Electric_Co.	53 Oklahoma_Gas_&_Electric_Co.
15 Cleveland_Electric_Illuminating_Co.	54 Pacific_Gas_&_Electric_Co.
16 Columbus_Southern_Power_Co.	55 PacifiCorp
17 Commonwealth_Edison_Co.	56 PECO_Energy_Co.
18 Consolidated Edison Co. of New York, Inc.	57 PP&L,_Inc.
19 Dayton_Power_&_Light_Co.	58 Potomac_Edison_Co.
20 Delmarva_Power_&_Light_Co.	59 Potomac_Electric_Power_Co.
21 Detroit_Edison_Co.	60 PSC_of_Colorado
22 Duke_Energy_Corp.	61 PSC_of_New_Hampshire
23 Duquesne_Light_Co.	62 PSC_of_New_Mexico
24 Entergy_Arkansas,_Inc.	63 PSI_Energy,_Inc.
25 Entergy_Gulf_States,_Inc.	64 Public_Service_Electric_&_Gas_Co.
26 Entergy_Louisiana,_Inc.	65 Rochester_Gas_&_Electric_Corp.
27 Entergy_Mississippi,_Inc.	66 San_Diego_Gas_&_Electric_Co.
28 Entergy_New_Orleans,_Inc.	67 South_Carolina_Electric_&_Gas_Co.
29 Florida_Power_&_Light_Co.	68 Southern_California_Edison_Co.
30 Florida_Power_Corp.	69 Southwestern_Electric_Power_Co.
31 Georgia_Power_Co.	70 Southwestern_Public_Service_Co.
32 Gulf_Power_Co.	71 Tampa_Electric_Co.
33 Houston_Lighting_&_Power_Co.	72 Texas_Utilities_Electric_Co.
34 Illinois_Power_Co.	73 United_Illuminating_Co.
35 Indiana_Michigan_Power_Co.	74 Virginia_Electric_&_Power_Co._(Virginia_Power)
36 Indianapolis_Power_&_Light_Co.	75 West_Penn_Power_Co.
37 Interstate_Power_Co.	76 Wisconsin_Electric_Power_Co.
38 Kansas_City_Power_&_Light_Co.	77 Wisconsin_Public_Service_Corp.
39 Kentucky_Utilities_Co.	

Table 2. Optimal directions for all Models

Across-the-sample posterior means and standard deviations, in parentheses, of optimal directions.

model	K	L	E	R	IC	Sulfur	CO_2	SO_2	NO_x
I	-0.876 (0.172)	0.485 (0.135)	0.335 (0.175)			-0.56 (0.175)	0.440 (0.125)	0.210 (0.151)	0.362 (0.155)
II	-0.85 (0.112)	0.594 (0.082)	0.526 (0.157)						
III	-0.652 (0.241)	0.711 (0.093)	0.820 (0.225)				0.350 (0.171)		
IV	-0.593 (0.191)	0.354 (0.084)	0.655 (0.176)	-0.662 (0.172)	-0.325 (0.125)				
V	-0.451 (0.112)	0.597 (0.095)	0.437 (0.167)	-0.512 (0.142)	0.620 (0.201)	-0.433 (0.154)	0.470 (0.150)	-0.651 (0.255)	-0.722 (0.254)

Notes: I, II and III are types of cost-minimization models, while IV and V are types of profit-maximization models.

Figure 1: Optimal Posterior Directions -- Model I
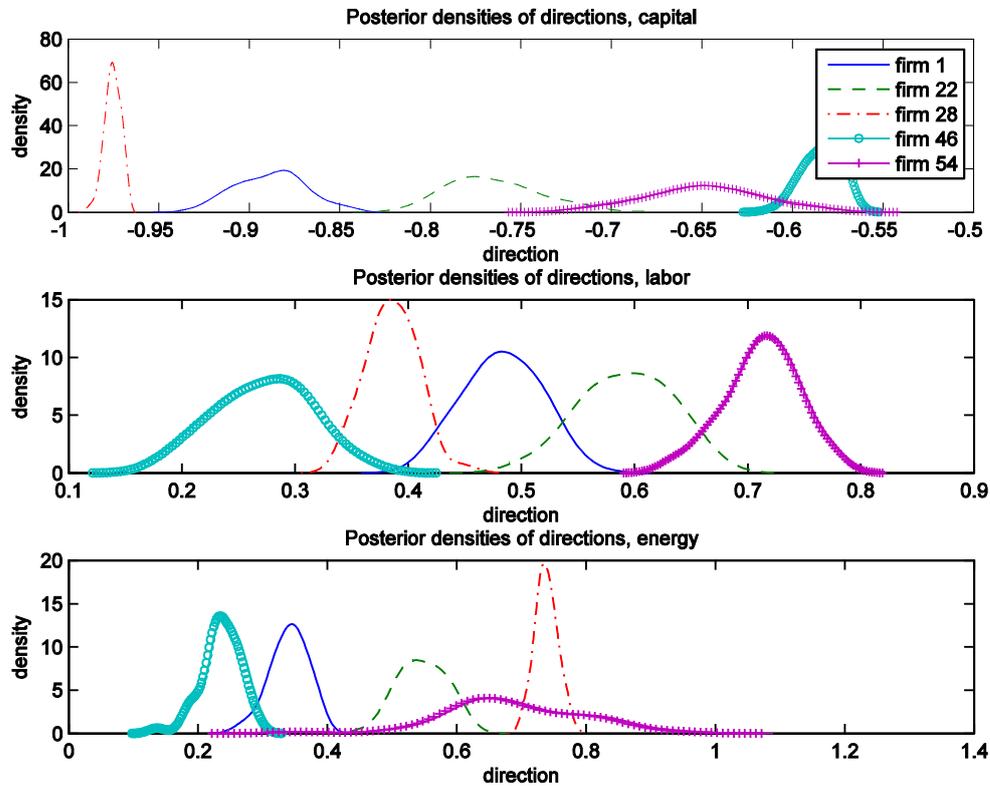
**Figure 2: Optimal Posterior Directions -- Model**

Posterior densities of directions, capital

Posterior densities of directions, labor

Posterior densities of directions, energy
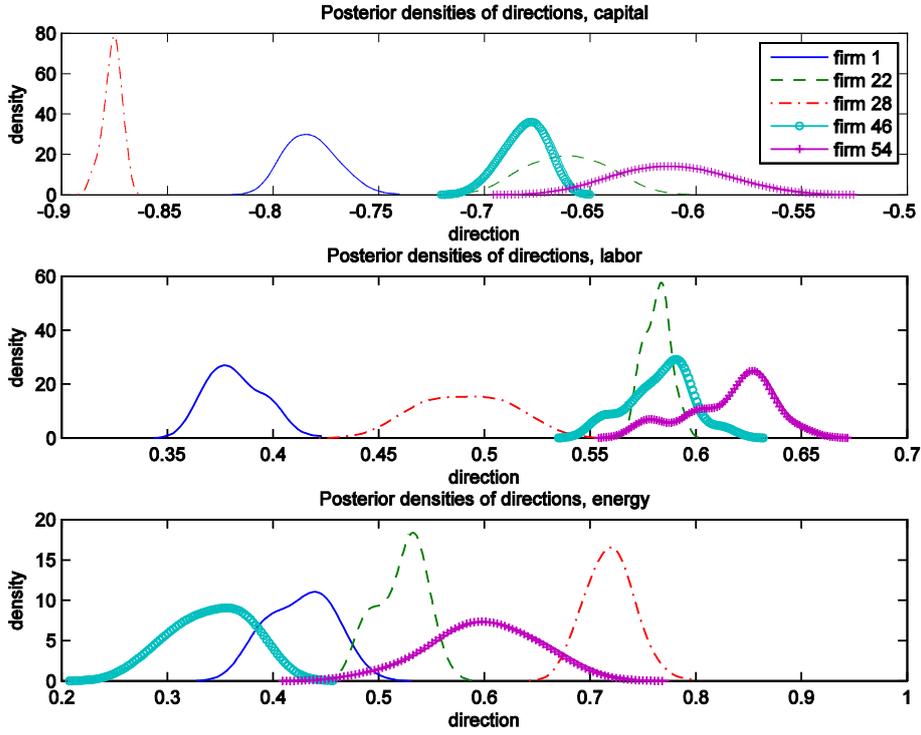
Figure 3: Optimal Posterior Directions -- Model

Posterior densities of directions, capital

Posterior densities of directions, labor

Posterior densities of directions, energy
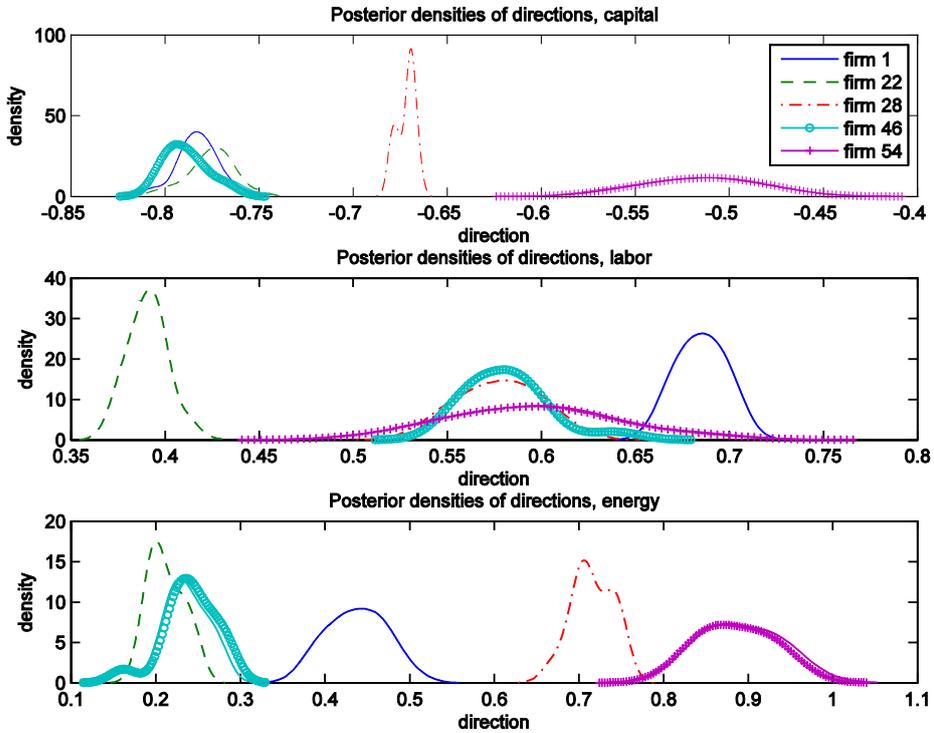
Figure 4: Optimal Posterior Directions -- Model IV
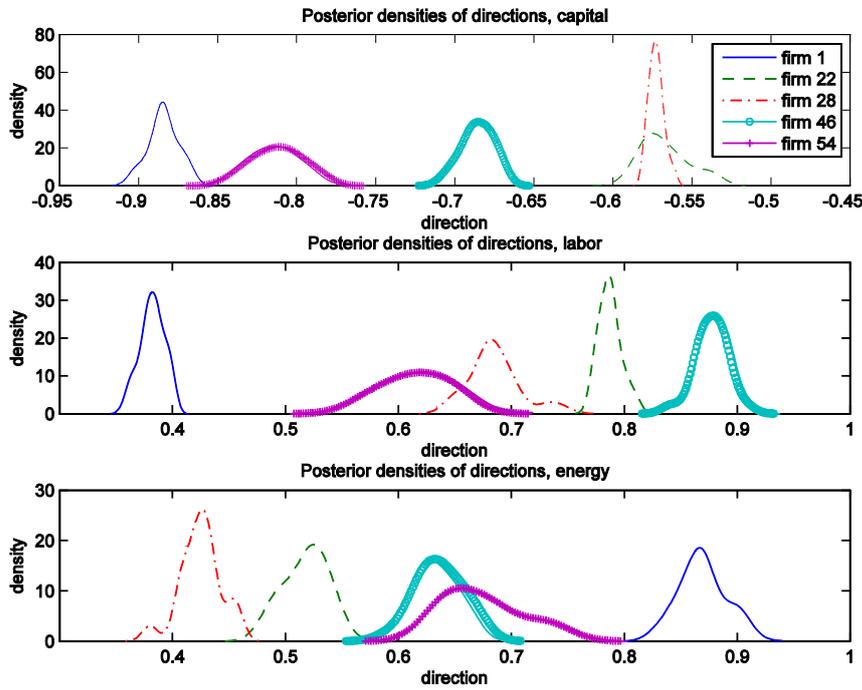
Figure 5: Optimal Posterior Directions -- Model V
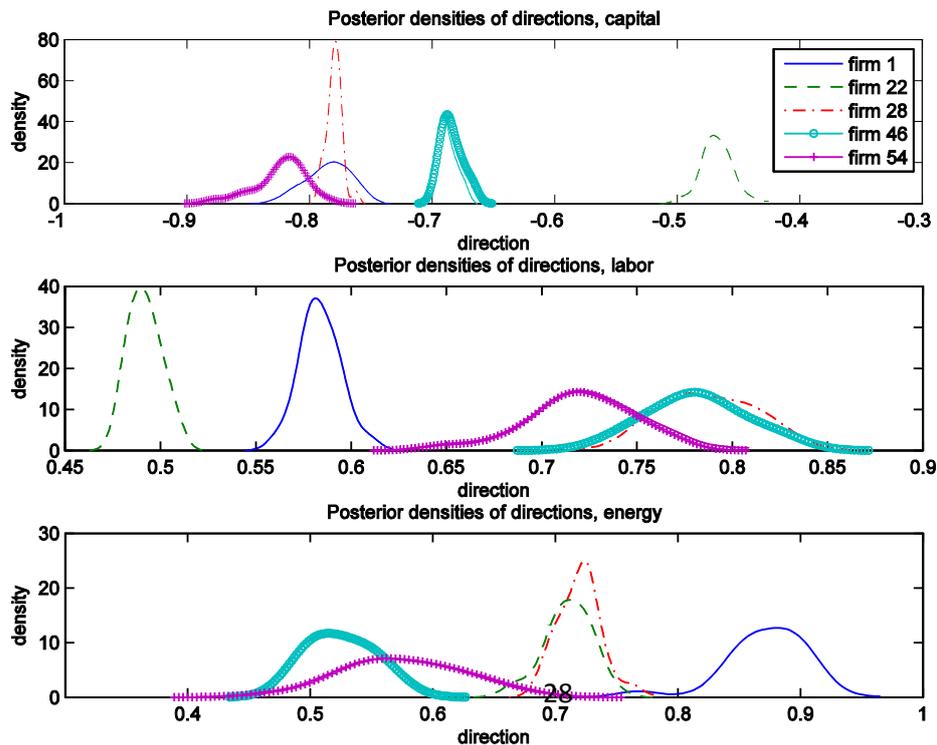
Figure 6a: Optimal Posterior Directions-Inputs and Good Outputs --Models IV and V
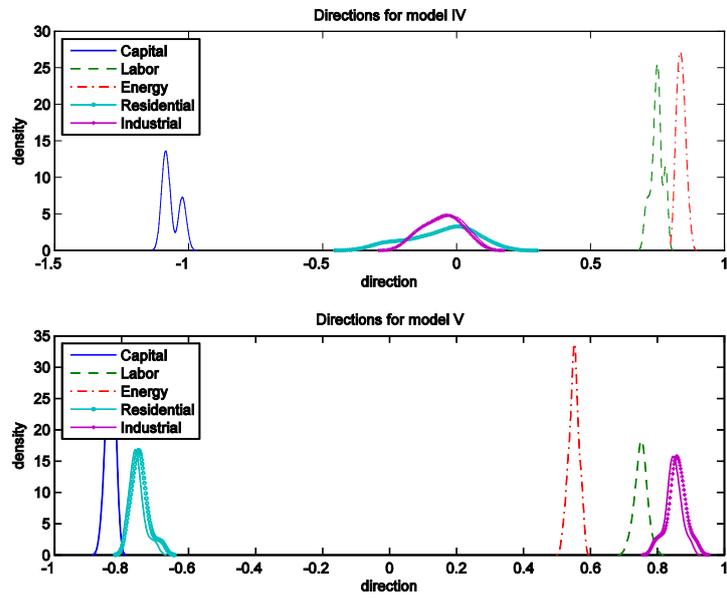
Figure 6b: Optimal Posterior Directions-Bad Input and Bad Outputs for Model V
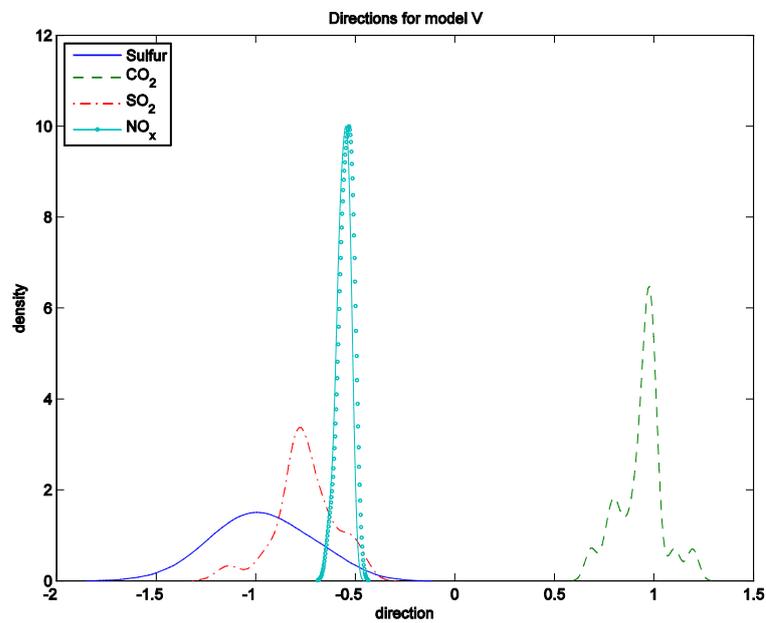
Table 3. Percent Changes in Inputs and Outputs (decimals)--All Models

Across-the-sample posterior means and standard deviations, in parentheses, of changes in inputs and outputs.

model	K	L	E	R	IC	Sulfur	CO$_2$	SO$_2$	NO$_x$
I	-0.051 (0.022)	0.027 (0.013)	0.021 (0.008)			-0.025 (0.009)	0.035 (0.009)	0.027 (0.009)	0.038 (0.012)
II	-0.055 (0.011)	0.061 (0.015)	0.048 (0.014)						
III	-0.043 (0.015)	0.078 (0.022)	0.061 (0.013)				0.033 (0.015)		
IV	-0.051 (0.022)	0.038 (0.015)	0.073 (0.016)	-0.057 (0.009)	-0.028 (0.008)				
V	-0.043 (0.015)	0.060 (0.021)	0.044 (0.012)	-0.050 (0.008)	0.059 (0.013)	-0.044 (0.014)	0.036 (0.009)	-0.065 (0.015)	-0.061 (0.025)

Notes : I, II and III are types of cost- minimization models, while IV and V are types of profit-maximization models.

Figure 7a: Percent Changes in Inputs and Good Outputs for Models IV and V
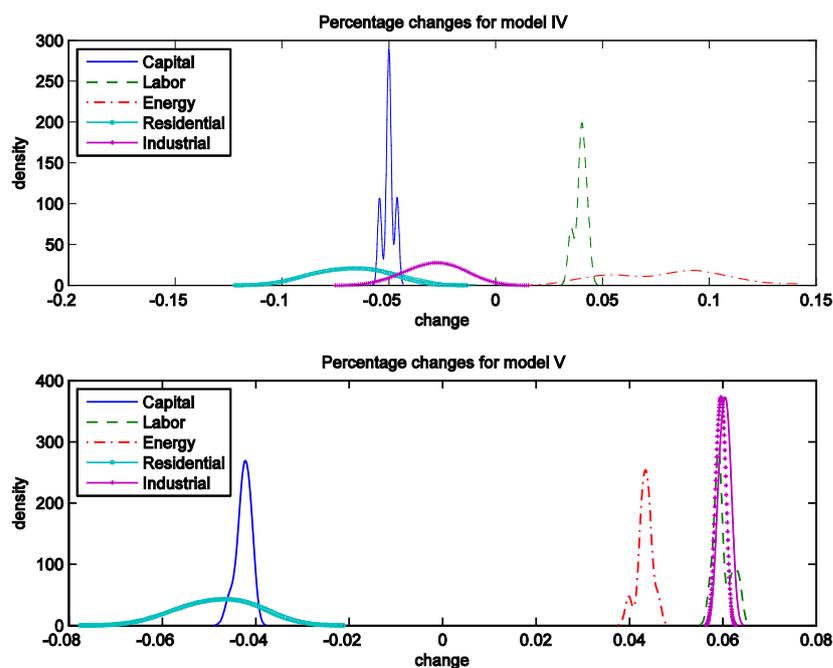
Figure 7b: Percent Changes in the Bad Input and Bad Outputs for Model V
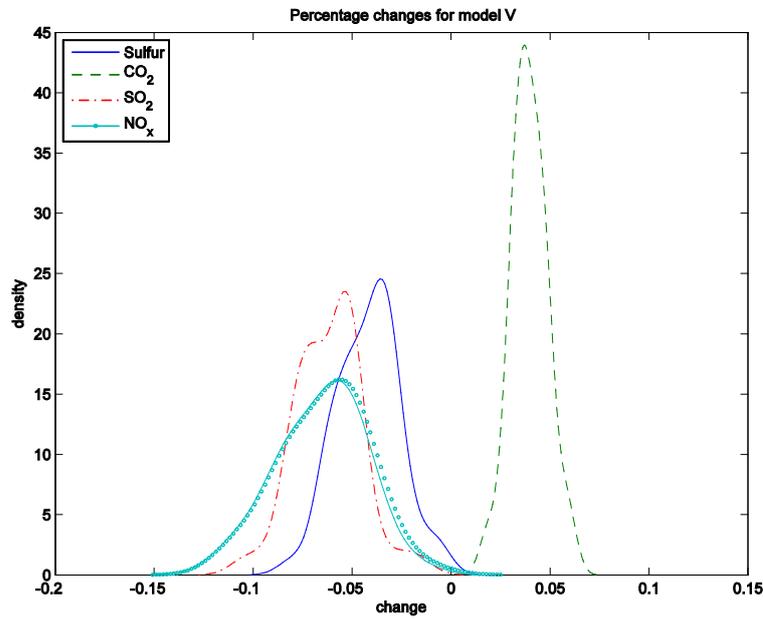
**Figure 8: Technical Efficiency for Models IV and V**
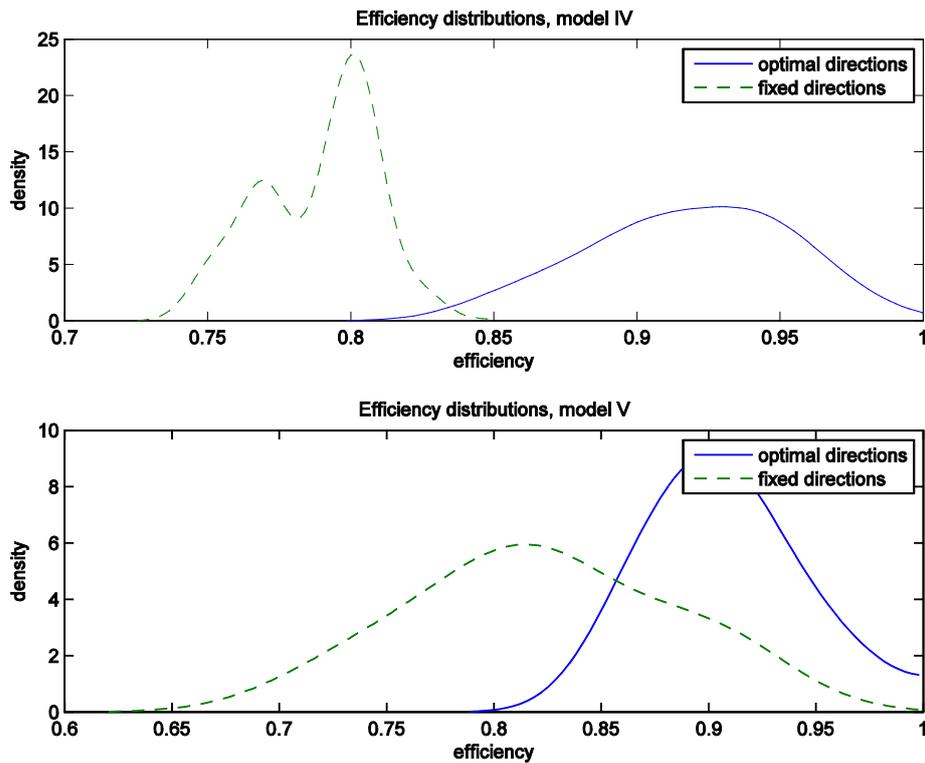
## Figure 9: TC for Models IV and V
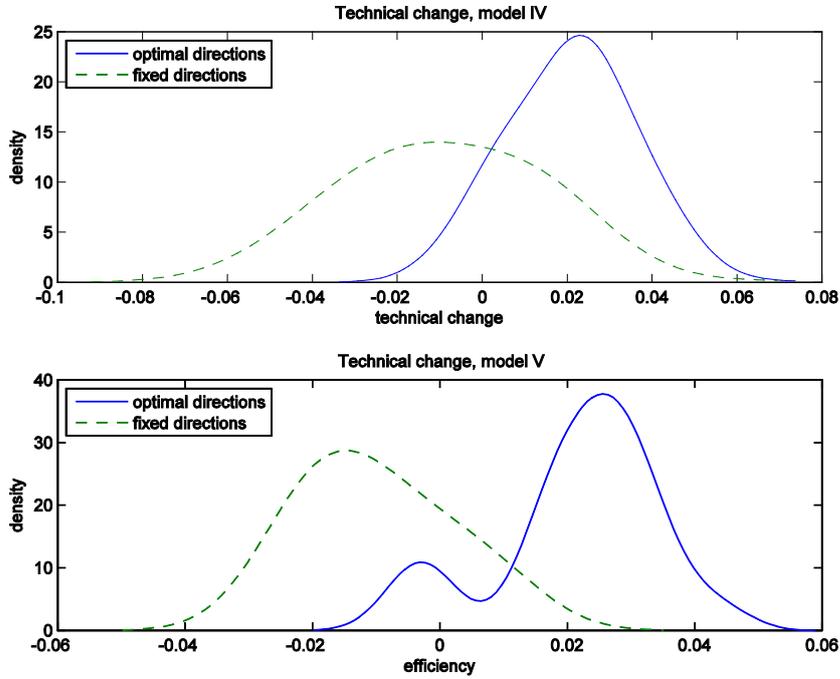
Figure 10: PC for Models IV and V
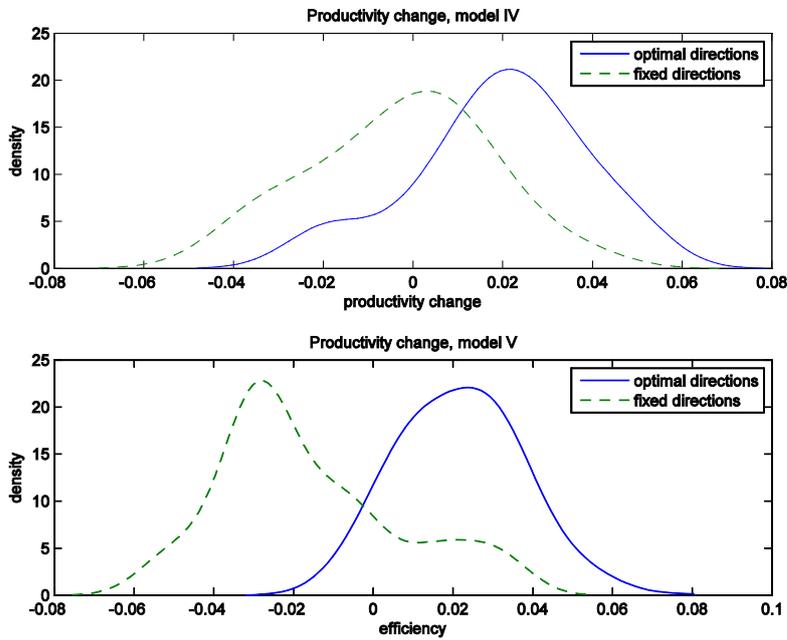
Table 4. MCMC convergence diagnostics

model	Max acf at lag 1	Max acf at lag 5	Max acf at lag 20	99% values GCD	99% values RNE
I	0.515	0.333	0.110	-1.712, 1.917	0.225, 0.615
II	0.412	0.312	0.171	-1.652, 1.815	0.171, 0.415
III	0.488	0.251	0.180	-1.782, 1.915	0.120, 0.551
IV	0.610	0.410	0.095	-1.601, 1.816	0.231, 0.415
V	0.625	0.425	0.122	-2.151, 2.017	0.171, 0.655

Notes: While acf is the autocorrelation function, GCD is Geweke's convergence diagnostic, and RNE is the relative numerical efficiency. The intervals are constructed by ranking the values of these statistics calculated across all parameters and then constructing a 99% interval around the median.

Figure 11: Sensitivity Analysis to Prior Parameters for Directions
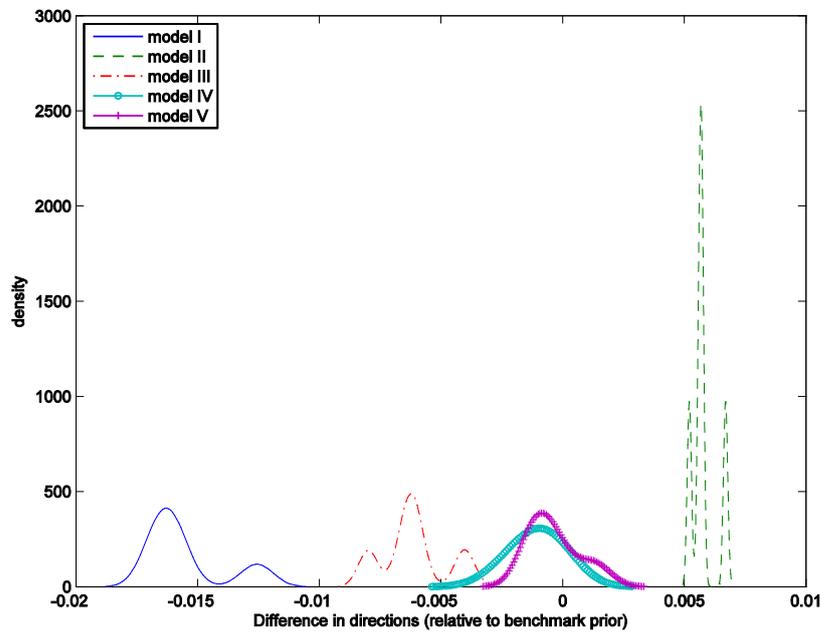
References

Agee, M.D., S. E. Atkinson, and T. D. Crocker, 2012, Child Maturation, Time-Invariant, and Time-Varying Inputs: their Interaction in the Production of Child Human Capital, *Journal of Productivity Analysis* 35, 29-44.

Atkinson, S. E. and J. H. Dorfman, 2005, Bayesian measurement of productivity and efficiency in the presence of undesirable outputs: crediting electric utilities for reducing air pollution, *Journal of Econometrics* 126, 445-468.

Averch, H. and L. L. Johnson, 1962, The Behavior of the Firm Under Regulatory Constraint, *American Economic Review*, 52, 1052–1069.

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., Y. Chung, and R. Färe, 1998, Profit, Directional Distance Functions, and Nerlovian Efficiency, *Journal of Optimization Theory and Applications*}} 98(2), 351-364.

Christensen, L.R. and D. W. Jorgenson, 1970, U.S. real product and real factor input, 1928--1967, *Review of Income and Wealth* 16, 19-50.

Färe, Rolf and S. Grosskopf, 2000, Theory and Application of Directional Distance Functions*, Journal of Productivity Analysis* 13(2), 93-103.

Färe, R., S. Grosskopf, and Y. H. Chung, 1997, Productivity and Undesirable Outputs: A Directional Distance Function Approach, *Journal of Environmental Management*, 51, 229-240.

Färe, R., S. Grosskopf, D. Noh, and W. Weber, 2005, Characteristics of a Polluting Technology: Theory and Practice, *Journal of Econometrics* 126, 469-492.

Fare, R., S. Grosskopf, G. Whittaker, 2013, Directional output distance functions: endogenous directions based on exogenous normalization constraints, *Journal of Productivity Analysis* 4, 267-269.

Feng, G. and A. Serletis, 2014, Undesirable outputs and a primal divisia productivity index based on the directional output distance function, forthcoming, *Journal of Econometrics*.

Federal Energy Regulatory Commission *Form 1: Annual Report of Major Electric Utilities*, various years.

Fernandez C., G. Koop and M.F.J. Steel, 2005, Alternative Efficiency Measures for Multiple-Output Production, *Journal of Econometrics* 126, 411-444.

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

Grosskopf, S. 2003, Some Remarks on Productivity and its Decompositions, *Journal of Productivity Analysis* 20, 459-474.

Hudgins, L. B. and D. Primont, 2007, Derivative Properties of Directional Technology Distance Functions, in Rolf Färe, S. Grosskopf, and D. Primont, eds., in *Aggregation, Efficiency, and Measurement,* Springer.

Moody's Public Utility Manual, Moody's Investors Service, New York, various years.

Pittman, R. W., 1983, Multilateral productivity Comparisons with Undesirable Outputs, *Economic Journal* 93, 883-891.

Rubin, C.B., 1988, Using the SIR algorithm to simulate posterior distributions (with discussion), in *Bayesian Statistics 3*, J.M. Bernardo, M.H. DeGroot, D.V. Lindley, and A.F.M Smith, eds, Oxford: Oxford University Press, pp. 395-402.

Smith, A.F.M., and A.E. Gelfand, 1992, Bayesian statistics without tears: a sampling-resampling procedure, *American Statistician* 46, 84-88.

Tsionas, E. G., E. Malikov, and S. C. Kumbhakar, 2014, Productivity Index based on the Directional Output Distance Function with Undesirable Outputs: Direction, Normalization and the Jacobian, working paper, Department of Economics, Athens University of Economics and Business, Athens, Greece.

U.S. Federal Power Commission, Steam Electric Plant Construction Cost and Annual Production Expenses}, Washington, D.C., annually.

U.S. Federal Power Commission, *Statistics of Privately Owned Electric Utilities in the U.S.--Classes A and B*, Washington D.C., annually.

U.S. Federal Power Commission, *Performance Profiles -- Private Electric Utilities in the United States: 1963--70*, Washington, D.C., annually.

Vardanyan, M. and Noh, D.W., 2006, Approximating pollution abatement costs via alternative specifications of a multi-output production technology: a case of the U.S. electric utility industry, *Journal of Environmental Management* 80, 177-190.

Yaisawarng, S. and J. D. Klein, 1994, The Effects of Sulfur Dioxide Controls on Productivity Change in the U.S. Electric Power Industry, *Review of Economics and Statistics* 76, 447-60,