Accepted Manuscript

Statistical inference in efficient production with bad inputs and outputs using latent prices and optimal directions

Scott E. Atkinson, Daniel Primont, Mike G. Tsionas

PII: S0304-4076(18)30016-2
DOI: https://doi.org/10.1016/j.jeconom.2017.12.009
Reference: ECONOM 4471

To appear in: *Journal of Econometrics*

Received date : 8 December 2016
Revised date : 13 July 2017
Accepted date : 3 December 2017

STATISTICAL INFERENCE IN EFFICIENT PRODUCTION WITH BAD INPUTS AND OUTPUTS USING LATENT PRICES AND OPTIMAL DIRECTIONS

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

Daniel Primont
Department of Economics
Southern Illinois University at Carbondale
email: primo@siu.edu

Mike G. Tsionas
Department of Economics
Lancaster University Management School
& Athens University of Economics and Business, Athens, Greece
LA1 4YX UK
email: m.tsionas@lancaster.ac.uk

January 23, 2018

Abstract

Researchers employ the directional distance function (DDF) to estimate multiple-input and multiple-output production, firm inefficiency, and productivity growth. We relax restrictive assumptions by computing optimal directions subject to profit maximization and cost minimization, correct for the potential endogeneity of inputs and outputs, estimate latent prices for bad outputs, measure firms' responses to shadow prices rather than actual prices, and introduce an unobserved productivity term into the DDF. For an unbalanced panel of U.S. electric utilities, a model assuming profit-maximization outperforms one assuming cost-minimization, while lagged productivity and energy price have the greatest effect on productivity.

JEL CODES: C11, C33, D24

KEY WORDS: Bayesian, directional distance, productivity, bad outputs, latent prices, efficiency, optimal directions, shadow prices

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

1 Introduction

As developed by Caves et al. (1982a, 1982b), the distance function (DF) has been widely used to estimate radial representations of frontier production technologies where firms employ multiple good inputs to produce multiple good outputs. The distance from a production frontier is a measure of the firm's technical efficiency (TE). The change in this measure over time is efficiency change (EC), while the shift in the frontier over time is technical change (TC). The sum of these two measures is productivity change (PC). 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.

One major shortcoming of the DF is that an entire set of inputs or an entire set of outputs must be scaled by the same factor. This becomes problematic when modelling the generation of electricity, since good inputs (capital, labor, and energy) and bad inputs (such as sulfur) produce good outputs (residential and industrial/commercial electricity) and bad outputs (pollutants). Using the DF, the researcher is not able to differentially credit the firm for simultaneously reducing bad outputs while increasing good outputs. In response, many authors have estimated an output DF and treated bad outputs like good inputs (holding both constant). However, this does not credit the firm for reducing bad outputs. Also, if bad inputs are consumed, no credit is given for their reduction.[1]

As an alternative, Chambers (1998) and Chambers et al. (1998) developed the directional distance function (DDF) which provides greater flexibility. It allows measurement of unique additive changes in each input and output through the calculation of different directions of movement for each to reach the production frontier. If non-zero directions are used to change only inputs (outputs), the DDF is input- (output-) oriented. When non-zero directions are used to change all inputs and outputs, the DDF is technology-oriented.

Despite the greater flexibility of the DDF, researchers typically impose three overly-restrictive assumptions. First, the researcher usually specifies arbitrary directions of movement of current firm production toward the frontier to measure inefficiency.[2] How-

[1] A bad input like sulfur would be consumed only when it is organically bound to the coal and oil which are burned to generate electricity. To our knowledge, only Yaisawarng and Klein (1994) include fuel sulfur content and sulfur dioxide emissions in a study of electric utility production.

[2] For example, assuming fixed directions, Färe et al. (2005) estimate an output DDF for electric utilities involving good inputs, a good output, and a bad output.

ever, different directions of movement toward the frontier will generate different measures of inefficiency. Three Data Envelopment Analysis (DEA) studies seek to avoid arbitrary assignment of directions by using linear programming methods to choose directions that *maximize* the measured distance (i.e., technical inefficiency) of the firm relative to a DDF. The first, by Färe et al. (2013), considers only good inputs and good outputs. The second, by Hampf and Krüger (2015), extends this analysis by including bad outputs. The stated goal of the third paper, by Zofio et al. (2013), is to compute optimal directions consistent with a firm's profit-maximization (PM) position on a DDF. They *assume* that firms are *currently* profit-maximizers and then measure the maximum distance from the current position. However, to measure the technology and productivity at the PM position, one must estimate the DDF jointly with the first-order conditions for PM. Since the latter are not included in their optimization model, the estimated directions cannot be consistent with PM. In this paper we estimate these conditions jointly with the DDF and compute directions consistent with PM, which we term "optimal-PM" directions.

Our approach follows Chambers (1998), who formulates a PM problem which includes a technology-oriented DDF (to measure the distance from the production frontier), and derives the first-order price equations for good inputs and outputs. In order to compute optimal-PM directions, Atkinson and Tsionas (2016) (AT) estimate the DDF jointly with the first-order price equations for only good inputs and good outputs, since the prices of bad outputs and bad inputs are missing. A complete set of utility-specific pollution permit prices (shadow prices for bad outputs) for the years of our sample data does not exist. As explained below, the prices of coal and oil include rebates for greater amounts of the bad input, sulfur. However, data is not publicly available to compute an hedonic price for sulfur.[3] We generalize AT by assuming a data generating process for latent prices of regulated bad outputs. These latent prices replace missing actual prices, allowing us to add the first-order price equations for regulated bad outputs to the AT system.

The second restrictive assumption of many DDF models is that all input and output quantities are exogenous. Highly-influential papers by Olley and Pakes (1996) (OP) and Levinsohn and Petrin (2003)(LP) consider the problem of estimating productivity in the

[3]In the more typical industry study, prices of all inputs are missing and our methodology can be employed to estimate their first-order price equations having generated their estimated latent prices.

presence of endogenous inputs using panel data. Both papers estimate a single-output Cobb-Douglas production function with a two-component random error term. The first component is firm- and time-varying productivity that is unobserved by the econometrician but observed, at least in part, by the firm. Since the firm takes productivity into account to some degree in choosing its inputs, endogeneity results. The second random component is an idiosyncratic error that is assumed to be uncorrelated with the explanatory variables and the productivity component. With the OP approach, the econometrician proxies for the unobserved productivity component with a potentially observable function. To obtain this function, OP first specify that investment is a monotonic function of productivity for a given level of capital and vintage. They then invert this function to obtain the productivity component as a proxy function of capital, investment, and vintage. Following OP, LP replace investment with materials and solve for the productivity component as a proxy function of capital, materials, and vintage. Productivity is assumed to follow a first-order Markov process. After discussing the modification of OP and LP by Ackerberg, Caves, and Frazer (2015) (regarding when the firm chooses labor), Wooldridge (2009) provides the exact set of moment conditions required to identify each of these models, where instruments are subsets of current and lagged inputs. However, as Griliches and Mairesse (1998) stress, if the econometrician incorrectly specifies the productivity function, some degree of endogeneity remains. Both OP and LP recognize the possible invalidity of their instruments as well as the typical validity but unavailability of input and output prices as instruments.

In this paper, we avoid assuming that inputs are exogenous for electric utilities. In our sample, they vary input choice over time and these choices are arguably correlated with the idiosyncratic error term, when one misspecifies the proxy equation for productivity. This results in the endogeneity of input quantities. Such a result potentially applies to all input quantities with a cost-minimization (CM) model and to all input and output quantities with a PM model. Instead, we utilize the prices of good inputs and good outputs in our instrument set, since they are arguably exogenous. Utilities are price takers in input markets, since these markets are national (due to trans-continental oil and natural gas pipelines, trans-continental rail lines hauling coal and oil, and national mobility of labor and capital). Regulated utilities, which comprise the vast majority of our sample, face output prices that are set by regulatory commissions. The smaller number of restructured utilities face market-determined prices for good inputs and outputs.[4] Thus, for both types of firms, we employ input and output prices rather than input quantities in our instrument set.

The third restrictive assumption with all previous DDF models is that actual prices equal shadow (perceived) prices for the firm.[5] If the two sets of prices differ, the researcher must calculate optimal directions using shadow prices. Previous papers have developed the methodology to estimate shadow prices for profit, cost, and distance functions as summarized in Kumbhakar and Lovell (2000). However, our paper is the first to estimate shadow prices using a DDF and the first-order price equations from PM. We identify shadow prices by including input and firm-specific price inefficiency parameters in these equations. These parameters are estimated jointly with optimal-PM directions.

In addition, this paper is the first to estimate a model free of these three restrictive assumptions and, at the same time, explain the sources of firm productivity, without resorting to inconsistent two-step methods. Typically the two steps are: 1) regress output on a set of inputs and 2) regress the residuals on a set of explanatory variables that were omitted from the first step. The two sets of variables must be uncorrelated to avoid a potentially substantial bias.[6] We avoid this improbable requirement by employing an unrestricted profit function from which we derive productivity as an estimable function of lagged productivity, profits, prices of inputs and outputs, vintage, and time. We include this measure of productivity as an input in the DDF. This enables us to compute the partial elasticities of productivity with respect to its arguments and decompose productivity growth.

We apply our methodology to an unbalanced panel of U.S. electric utilities. This sample significantly expands the AT data set by 80% to include years when a number of utilities were restructured. We report posterior densities for optimal directions, TE, EC, TC, PC, the resource implications of price inefficiency, and the sources of PC.

[4]The goal of deregulation was to increase competition, yielding greater TE, productivity growth, and price efficiency. On the production frontier, the profit-maximizing firm achieves price efficiency when the price of each input equals the value of its marginal product. The cost-minimizing firm achieves allocative efficiency when ratios of input prices equal ratios of their marginal products.

[5]Reasons for deviations of shadow from actual prices include tax write-offs, rate-of-return regulation, and constraints imposed by regulatory agencies or labor unions.

[6]See Wang and Schmidt (2002) for details on Monte Carlo experiments indicating substantial potential bias in both steps.

2 The Directional Distance Function

2.1 Computing Optimal Directions

We assume a firm production technology that combines good inputs, $\mathbf{x} = (x_1, \ldots, x_N) \in R_+^N$, and bad inputs, $\tilde{\mathbf{x}} = (\tilde{x}_1, \ldots, \tilde{x}_J) \in R_+^J$, to produce good outputs, $\mathbf{y} = (y_1, \ldots, y_M) \in R_+^M$, and bad outputs, $\tilde{\mathbf{y}} = (\tilde{y}_1, \ldots, \tilde{y}_L) \in R_+^L$. A firm with vintage, τ, productivity shock, ω, at time $t(t = 1, \ldots, T)$, has production technology

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

Let $\boldsymbol{g} = (\boldsymbol{g}_x, \boldsymbol{g}_{\tilde{x}}, \boldsymbol{g}_y, \boldsymbol{g}_{\tilde{y}})$ be a direction vector. Typically researchers assume that $(\boldsymbol{g}_x, \boldsymbol{g}_{\tilde{x}}, \boldsymbol{g}_y, \boldsymbol{g}_{\tilde{y}}) = (-\mathbf{1}, -\mathbf{1}, \mathbf{1}, -\mathbf{1})$. Following Chambers (1998), we define the technology DDF as

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

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

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

D1. Translation Property:

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

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

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

D3. Concavity:

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

D4. Non-negativity:

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

Equation (3) says that the technology DDF will satisfy the translation property. For example, increasing $\mathbf{y}$ and decreasing $\mathbf{x}, \tilde{\mathbf{x}}$, and $\tilde{\mathbf{y}}$ by α, each multiplied by their direction, will result in a decrease in the technology DDF by α. This is analogous to the property of linear homogeneity with a Shephard distance function. Equation (4) indicates that scaling each direction by λ will scale the technology DDF by λ^{-1}. Equation (5) imposes concavity of the technology DDF. Finally, equation (6) requires that the technology DDF function be non-negative, which is easily imposed after estimation.

For the following properties we let "S" represent the assumption of strong disposability of all inputs and outputs:

D5–S. Good Input Monotonicity: We first assume that good inputs are strongly disposable. In Appendix A.1 we show that this implies[7]

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

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

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

Following the proof of D5–S, assuming that bad inputs are strongly disposable, we can determine

D7–S. Bad Input Monotonicity:

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

Again following the proof of D5–S, assuming that bad outputs are strongly disposable, we can determine

D8–S. Bad Output Monotonicity:

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

However, researchers typically do not assume that bads are strongly disposable. Instead, most investigators maintain that bad inputs and bad outputs are weakly disposable with good inputs and outputs. The weak disposability of bad inputs results from their organic combination with good inputs, so that a reduction in bad inputs implies a reduction in good inputs, outputs held constant. Likewise, bad outputs are weakly disposable since to reduce them, we must divert resources from the production of good outputs, holding inputs constant. Assuming weak disposability, we cannot determine

[7]The paper with all Appendices is stored at *people.terry.uga.edu/directory/profile/atknsn/* and at RePec.

the global monotonicity property of bad inputs and bad outputs as in D7–S and D8–S. However, as shown shortly, if we assume weak disposability, we can determine local monotonicity conditions given that we also assume PM and know the signs of the prices of bad inputs and bad outputs. If instead we assume CM, the firm is subject to only the first-order conditions for inputs, so that the monotonicity conditions for good and bad inputs are the same as with the PM model.

A number of studies have modeled electric utilities assuming PM. See Atkinson and Halvorsen (1976, 1980) and Cowing (1978). Other researchers have assumed CM subject to a set of regulatory constraints on earned rates of return and a requirement to satisfy all demand at a given price (that is, output is taken as given). However, output may be endogenous as with an ex ante cost function.[8] Further, if these regulatory constraints are not binding, utilities may maximize profits. Fowlie (2010) provides evidence of this by showing that many regulated utilities earn allowed rates of return on capital that considerably exceed the market rate of return, indicating that constraints on profits may not be binding and output may be endogenous. These results indicate that a PM model may be more appropriate than a CM model. Thus, we focus on the PM model and compare the accuracy of its results to those of the CM model.

Temporarily suppressing all the arguments of $\overrightarrow{D}_{\mathcal{T}}(\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y}, \tilde{\mathbf{y}}; \boldsymbol{g}, \omega, \tau, t)$, we follow Chambers (1998) and assume that a firm maximizes profits, π, by choosing values of $\mathbf{x}, \mathbf{y}, \tilde{\mathbf{x}}, \tilde{\mathbf{y}}$ to solve

$$\sup \left\{\mathbf{p}_{\mathrm{y}}(\mathbf{y} + \overrightarrow{D}_{\mathcal{T}}\boldsymbol{g}_{\mathrm{y}}) - \mathbf{p}_{\tilde{\mathrm{y}}}(\tilde{\mathbf{y}} + \overrightarrow{D}_{\mathcal{T}}\boldsymbol{g}_{\tilde{\mathrm{y}}}) - \mathbf{p}_{\mathrm{x}}(\mathbf{x} + \overrightarrow{D}_{\mathcal{T}}\boldsymbol{g}_{\mathrm{x}}) - \mathbf{p}_{\tilde{\mathrm{x}}}(\tilde{\mathbf{x}} + \overrightarrow{D}_{\mathcal{T}}\boldsymbol{g}_{\tilde{\mathrm{x}}})\right\}, \tag{11}$$

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

Typically (as indicated above) the DDF is estimated without the first-order conditions for PM, where one has assumed a set of *a priori* fixed directions. However, in this paper we assume that the firm chooses $(\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y}, \tilde{\mathbf{y}})$ subject to the first-order conditions for PM and we estimate optimal-PM directions consistent with these conditions.[9] The

[8]See Pope and Just (1996) regarding identification and estimation of an ex ante cost function.

[9]The CM model is obtained by using only the first-order conditions for input prices and assuming that output is given.

first-order conditions are:

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

where

$$\varrho(\mathbf{p},\boldsymbol{g}) = [\sum_m p_m g_m - \sum_n p_n g_n - \sum_j p_j g_j - \sum_l p_l g_l]. \tag{16}$$

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

$$\mathcal{L} = \mathbf{p}_{\mathrm{y}}\mathbf{y} - \mathbf{p}_{\mathrm{x}}\mathbf{x} - \mathbf{p}_{\tilde{\mathrm{y}}}\tilde{\mathbf{y}} - \mathbf{p}_{\tilde{\mathrm{x}}}\tilde{\mathbf{x}} + \varrho\overrightarrow{D}_{\mathcal{T}}(\mathbf{x},\tilde{\mathbf{x}},\mathbf{y},\tilde{\mathbf{y}};\boldsymbol{g},\omega,\tau,t), \tag{17}$$

in order to obtain this interpretation of ϱ.

All prices are assumed to be non-negative except for the price of bad inputs, which is non-positive. The prices of good outputs and the prices of good inputs are non-negative by definition. The price of bad outputs is positive, since the firm must pay a fine or buy emission permits for additional production of bad outputs. When a bad input is organically bound to a good input, the price of the bad input is negative, since the firm must be compensated for utilizing it.

For the following two properties, "W" indicates weakly disposable. Assuming weak disposability of bad inputs and that $\mathbf{p}_{\tilde{x}} \leq 0$, from (14) we obtain locally:

D7–W. Bad Input Monotonicity:

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

Assuming weak disposability of bad outputs and that $\mathbf{p}_{\tilde{y}} \geq 0$ from (15) we obtain locally:

D8–W. Bad Output Monotonicity:

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

In this manner we maintain PM and use assumptions about the prices of bad inputs and bad outputs to locally restrict the range of the partial derivatives in D7–W and D8–W. For the CM model, only the monotonicity conditions for good and bad inputs apply. They are the same as with the PM model.

Rather than assume fixed directions, in this paper we compute optimal directions that are consistent with PM and CM. In Fig. 1, a firm increases a good output and reduces a bad output by moving from z to z^*, the PM point. This is consistent with the traditional assumptions, where the directions for the good output and bad output are positive and negative, respectively. However, the signs of optimal directions may be quite different. In Fig. 1, if a firm moved from z to the point of PM at z', from (2) the directions $(\boldsymbol{g}_{\mathrm{y}}, \boldsymbol{g}_{\tilde{\mathrm{y}}})$ are both positive since the good output and the bad output would both increase. One can easily substitute $\tilde{\mathbf{x}}$ or $\mathbf{x}$ for $\tilde{\mathbf{y}}$ and obtain similar results. An electric utility subject to the first-order conditions for PM may maximize profits by increasing good outputs, increasing some inputs to produce the additional good outputs, decreasing regulated bad outputs, and increasing unregulated bad outputs due to the increase in good outputs. Thus, we impose no a priori sign restrictions on the optimal direction of any input or output.

The input and output DDF are easily obtained as special cases of the technology DDF. The output DDF changes good and bad outputs in the direction $(\mathbf{0}, \mathbf{0}, \boldsymbol{g}_{\mathrm{y}}, \boldsymbol{g}_{\tilde{\mathrm{y}}})$, for a given level of inputs in order to move to the frontier of $\mathcal{T}(\omega, \tau, t)$. The input DDF changes good and bad inputs in the direction $(\boldsymbol{g}_{\mathrm{x}}, \boldsymbol{g}_{\tilde{\mathrm{x}}}, \mathbf{0}, \mathbf{0})$, for a given level of good and bad outputs in order to move to the frontier of $\mathcal{T}(\omega, \tau, t)$.

3 Econometric Formulation

3.1 The Technology Directional Distance System and the Translation Restrictions

Assume that we have panel data for firm $i(i = 1, \ldots, \mathcal{N})$ in time period $t(t = 1, \ldots, T)$ on all inputs and outputs. We then formulate our technology DDF as a quadratic function of $\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y}, \tilde{\mathbf{y}}, \tau$, and t as:

$$\begin{aligned}
\overrightarrow{D}_{\mathcal{T}}(\mathbf{x}_{it}, \tilde{\mathbf{x}}_{it}, \mathbf{y}_{it}, \tilde{\mathbf{y}}_{it}; \tau_{it}, t) \quad = \quad & \sum_{n=1}^{N} \gamma_n (x_{n,it}) + \sum_{m=1}^{M} \gamma_m (y_{m,it}) + \sum_{j=1}^{J} \gamma_j (\tilde{x}_{j,it}) + \sum_{l=1}^{L} \gamma_l (\tilde{y}_{l,it}) \\
+ \; & \frac{1}{2} \sum_{n=1}^{N} \sum_{n'=1}^{N} \gamma_{nn'} (x_{n,it})(x_{n',it}) + \frac{1}{2} \sum_{j=1}^{J} \sum_{j'=1}^{J} \gamma_{jj'} (\tilde{x}_{j,it})(\tilde{x}_{j',it}) \\
+ \; & \frac{1}{2} \sum_{m=1}^{M} \sum_{m'=1}^{M} \gamma_{mm'} (y_{m,it})(y_{m',it}) + \frac{1}{2} \sum_{l=1}^{L} \sum_{l'=1}^{L} \gamma_{ll'} (\tilde{y}_{l,it})(\tilde{y}_{l',it})
\end{aligned}$$

$$
\begin{aligned}
&+ \sum_{j=1}^{J}\sum_{n=1}^{N}\gamma_{jn}(\tilde{x}_{j,it})(x_{n,it}) + \sum_{m=1}^{M}\sum_{n=1}^{N}\gamma_{mn}(y_{m,it})(x_{n,it}) \\
&+ \sum_{l=1}^{L}\sum_{n=1}^{N}\gamma_{ln}(\tilde{y}_{l,it})(x_{n,it}) + \sum_{m=1}^{M}\sum_{j=1}^{J}\gamma_{jm}(y_{m,it})(\tilde{x}_{j,it}) \\
&+ \sum_{l=1}^{L}\sum_{j=1}^{J}\gamma_{jl}(\tilde{y}_{l,it})(\tilde{x}_{j,it}) + \sum_{l=1}^{L}\sum_{m=1}^{M}\gamma_{lm}(\tilde{y}_{l,it})(y_{m,it}) \\
&+ \sum_{n=1}^{N}\gamma_{nt}(x_{n,it})\, d_t + \sum_{m=1}^{M}\gamma_{mt}(y_{m,it})\, d_t + \sum_{j=1}^{J}\gamma_{jt}(\tilde{x}_{j,it})\, d_t \\
&+ \sum_{l=1}^{L}\gamma_{lt}(\tilde{y}_{l,it})\, d_t \;+ \sum_{t=1}^{T}\gamma_t d_t + \sum_{n=1}^{N}\gamma_{n\tau}(x_{n,it})\,\tau_{it} \\
&+ \sum_{m=1}^{M}\gamma_{m\tau}(y_{m,it})\,\tau_{it} + \sum_{j=1}^{J}\gamma_{j\tau}(\tilde{x}_{j,it})\,\tau_{it} \\
&+ \sum_{l=1}^{L}\gamma_{l\tau}(\tilde{y}_{l,it})\,\tau_{it} \;+ \gamma_\tau \tau_{it}, \qquad (20)
\end{aligned}
$$

where d_t is a year dummy. We later accept the null that d_t and τ_{it} enter (20) linearly using a likelihood ratio test at the .05 level. Hence, we drop all interaction terms involving these variables from this equation.[10]

We can now restate the first-order conditions for PM, (12)-(15), in terms of the parameters of the quadratic DDF for each good input price equation as

$$
p_{n,it}/\varrho = \gamma_n + \sum_{n'=1}^{N}\gamma_{nn'}(x_{n',it}) + \sum_{j=1}^{J}\gamma_{jn}(\tilde{x}_{j,it}) + \sum_{m=1}^{M}\gamma_{mn}(y_{m,it}) + \sum_{l=1}^{L}\gamma_{ln}(\tilde{y}_{l,it}), \qquad (21)
$$

for each good output price equation as

$$
p_{m,it}/\varrho = -\Bigg[\gamma_m + \sum_{m'=1}^{M}\gamma_{mm'}(y_{m',it}) + \sum_{j=1}^{J}\gamma_{jm}(\tilde{x}_{j,it}) + \sum_{n=1}^{N}\gamma_{mn}(x_{n,it}) + \sum_{l=1}^{L}\gamma_{lm}(\tilde{y}_{l,it})\Bigg], \qquad (22)
$$

for each bad input price equation as

$$
p_{j,it}/\varrho = \gamma_j + \sum_{j'=1}^{J}\gamma_{jj'}(\tilde{x}_{j',it}) + \sum_{n=1}^{N}\gamma_{jn}(x_{n,it}) + \sum_{m=1}^{M}\gamma_{jm}(y_{m,it})
$$

[10] Färe and Lundberg (2005) prove that only two functional forms have a second-order Taylor series approximation interpretation of the DDF and satisfy the translation property. These are the logarithmic transcendental and the quadratic. Also see Chambers (1998) for further discussion of this point. We employ the quadratic since it is linear in the parameters.

$$+\sum_{l=1}^{L}\gamma_{jl}(\tilde{y}_{l,it}), \tag{23}$$

and for each bad output price equation as

$$p_{l,it}/\varrho = \gamma_l + \sum_{l'=1}^{L}\gamma_{ll'}(\tilde{y}_{l',it}) + \sum_{j=1}^{J}\gamma_{jl}(\tilde{x}_{j,it}) + \sum_{n=1}^{N}\gamma_{ln}(x_{n,it}) + \sum_{m=1}^{M}\gamma_{lm}(y_{m,it}). \tag{24}$$

The restrictions guaranteeing the translation property in (3) are imposed parametrically on (20) and on (21)-(24).[11] To derive these restrictions for an input, an output, and a technology DDF, we generalize Hudgins and Primont (2007) by adding bad inputs and bad outputs. To simplify notation, let $\tilde{\mathbf{z}} = (\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y}, \tilde{\mathbf{y}})$. First, assuming a quadratic functional form for the technology DDF:

$$\overrightarrow{D}_{\mathcal{T}}(\tilde{\mathbf{z}}) = \sum_{w=1}^{W}\gamma_w\tilde{z}_w + \sum_{w=1}^{W}\sum_{w'=1}^{W}\gamma_{ww'}\tilde{z}_w\tilde{z}_{w'}, \tag{25}$$

where $w = 1, \ldots, W$, $W = M + N + J + L$. To determine the appropriate parametric restrictions that guarantee the translation property incorporating $\boldsymbol{g}$, we note that the translation property requires that

$$\begin{aligned}\overrightarrow{D}_{\mathcal{T}}(\tilde{\mathbf{z}} + \alpha\boldsymbol{g}; \boldsymbol{g}) &= \sum_{w=1}^{W}\gamma_w(\tilde{z}_w + \alpha g_w) \\ &+ \sum_{w=1}^{W}\sum_{w'=1}^{W}\gamma_{ww'}(\tilde{z}_w + \alpha g_w)(\tilde{z}_{w'} + \alpha g_{w'}) \\ &= \sum_{w=1}^{W}\gamma_w\tilde{z}_w + \sum_{w=1}^{W}\sum_{w'=1}^{W}\gamma_{ww'}\tilde{z}_w\tilde{z}_{w'} - \alpha. \end{aligned} \tag{26}$$

3.2 The Stochastic Framework with Shadow and Latent Prices

3.2.1 Stochastically Imposing the Translation Property Restrictions

Generalizing Hudgins and Primont (2007), we derive the following parametric restrictions to stochastically impose the translation property in (26) for the technology

[11] An alternative to the quadratic is the logarithmic transcendental technology translation function which automatically satisfies the translation properties. See Chambers (1998) for more details. Note that we can approximate a DDF with a quadratic function, but not a translog. Using the former, one can impose the translation property, since original variables are used. However, one cannot impose the translation property on the latter.

DDF:

$$
\begin{aligned}
&\sum_{m=1}^{M} \gamma_m g_m + \sum_{l=1}^{L} \gamma_l g_l + \sum_{n=1}^{N} \gamma_n g_n + \sum_{j=1}^{J} \gamma_j g_j = \vartheta_o, \\
&\sum_{m=1}^{M} \gamma_{mn'} g_m + \sum_{l=1}^{L} \gamma_{ln'} g_l + \sum_{n=1}^{N} \gamma_{nn'} g_n + \sum_{j=1}^{J} \gamma_{jn'} g_j = \vartheta, \ \forall\, n', \\
&\sum_{m=1}^{M} \gamma_{mm'} g_m + \sum_{l=1}^{L} \gamma_{lm'} g_l + \sum_{n=1}^{N} \gamma_{m'n} g_n + \sum_{j=1}^{J} \gamma_{jm'} g_j = \vartheta, \ \forall\, m', \\
&\sum_{m=1}^{M} \gamma_{j'm} g_m + \sum_{l=1}^{L} \gamma_{j'l'} g_l + \sum_{n=1}^{N} \gamma_{j'n} g_n + \sum_{j=1}^{J} \gamma_{jj'} g_j = \vartheta, \ \forall\, j', \\
&\sum_{m=1}^{M} \gamma_{l'm} g_m + \sum_{l=1}^{L} \gamma_{ll'} g_l + \sum_{n=1}^{N} \gamma_{l'n} g_n + \sum_{j=1}^{J} \gamma_{jl'} g_j = \vartheta, \ \forall\, l',
\end{aligned}
\tag{27}
$$

where $\vartheta_o \sim \mathscr{N}\left(-1, c^2\right)$ and $\vartheta \sim \mathscr{N}\left(0, c^2\right)$, $\mathscr{N}$ is the normal density, and $c = 10^{-4}$ to keep the variance small around zero. The stochastic restrictions in (27) are, in fact, *semi-informative priors* placed upon $\boldsymbol{g}$ and $\boldsymbol{\gamma}$, a vector of all γ_w parameters. These parameters are estimated jointly, as explained below. In a more traditional, non-Bayesian approach, one would set $\vartheta_o = -1, \vartheta = 0$.

To obtain the corresponding restrictions for the input and output DDF models, simply eliminate the summation terms for the outputs and the inputs, respectively. For the technology DDF model, we have imposed symmetry for all the double-subscripted coefficients for all inputs and outputs. Note that the introduction of parameters to measure the direction for each input and output occurs *only* through the translation property restrictions.

3.2.2 Modeling ω

Assuming now that production is a function of inputs, outputs, vintage and a Hicks-neutral productivity shock, ω, the resulting profit function obtained from (11) is

$$\pi = \pi(\mathbf{p}_{\mathrm{y}}, \mathbf{p}_{\tilde{\mathrm{y}}}, \mathbf{p}_{\mathrm{x}}, \mathbf{p}_{\tilde{\mathrm{x}}}, \omega; \boldsymbol{g}, \tau, t) = \pi(\mathbf{p}, \omega; \boldsymbol{g}, \tau, t). \tag{28}$$

By definition, π is non-decreasing, monotonic in ω. Strengthening this condition to positive monotonicity of π in ω, we can write the inverse function for ω as $\omega = f(\mathbf{p}, \pi; \boldsymbol{g}, \tau, t)$.

Subject to (27) we specify a stochastic framework for (20) as:

$$0 = \overrightarrow{D}_{\mathcal{T}}(\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y}, \tilde{\mathbf{y}}; \boldsymbol{g}, \tau, t) + v_{it} + \omega_{it} - u_{it}, \tag{29}$$

where the stochastic part is comprised of an idiosyncratic i.i.d. term, v_{it}, which has zero mean, ω_{it}, and a one-sided component, u_{it}. While v_{it} reflects errors in optimization due to random events beyond the control of the firm (such as weather), u_{it} reflects firm-specific inefficiencies that may vary over time. We generalize the productivity component, ω_{it}, by including $\omega_{i,t-1}$, which is lagged ω_{it}, and utilizing $\mathbf{p}^*_{it}$, where the star indicates that latent prices replace missing prices:

$$\omega_{it} = f(\omega_{i,t-1}, \mathbf{p}^*_{it}, \pi_{it}; \boldsymbol{g}, \tau_{it}, t) + \varepsilon_{it,1} \tag{30}$$

and we must obtain an approximation to the unknown functional form $f(\cdot)$. We further specify that

$$log(u_{it}) = \gamma_1 + \gamma_2\omega_{it} + \gamma_3\omega_{i,t-1} + \gamma_4 \log u_{i,t-1} + \gamma_5 t + \mathbf{d}'_{it}\boldsymbol{\gamma}_o + \mathbf{Z}'_{i,t-1}\delta + \varepsilon_{it,2}, \tag{31}$$

where $\mathbf{d}_{it}$ denotes firm dummies, and $\mathbf{Z}_{i,t-1}$ contains lagged values of all inputs and outputs.

We now consider different approximations of (30). To obtain a translog-neural-network approximation, let $\mathbf{z}_{it} = (\omega_{i,t-1}, \mathbf{p}^*_{it}, \pi_{it}; \boldsymbol{g}, \tau_{it}, t)$. Then

$$f(\mathbf{z}_{it}) = a_o + \mathbf{a}'\mathbf{z}_{it} + \tfrac{1}{2}\mathbf{z}'_{it}\mathbf{A}\mathbf{z}_{it} + \sum_{g=1}^{G} \lambda_g \varphi\left(\mathbf{z}'_{it}\mathbf{b}_g\right), \tag{32}$$

where the activation function $\varphi(\kappa) = \frac{1}{1+\exp(-\kappa)}$, $-\infty < \kappa < \infty$. As a second alternative, we use a second-order approximation:

$$f(\mathbf{z}_{it}) = a_o + \mathbf{a}'\mathbf{z}_{it} + \tfrac{1}{2}\mathbf{z}'_{it}\mathbf{A}\mathbf{z}_{it}. \tag{33}$$

As a third alternative, we use the Fourier approximation:

$$\begin{gathered} f(\mathbf{z}_{it}) = a_o + \mathbf{a}'\mathbf{z}_{it} + \tfrac{1}{2}\mathbf{z}'_{it}\mathbf{A}\mathbf{z}_{it} + \\ \textstyle\sum_{\phi=1}^{\Phi} \left\{ u_{0\phi} + 2\sum_{\ell=1}^{\mathcal{J}} \left(u_{\ell\phi}\cos(\ell\mathbf{k}'_\phi\mathbf{z}_{it}) + v_{\ell\phi}\sin(\ell\mathbf{k}'_\phi\mathbf{z}_{it}) \right) \right\}, \end{gathered} \tag{34}$$

where $\mathbf{k}_\phi$ is a multi-index, Φ is a number determined by $dim(\mathbf{z}_{it})$, $\mathcal{J}$ is the order of the expansion, and $u_{0\phi}, u_{\ell\phi}, v_{\ell\phi}$ are unknown parameters. The multi-indices are constructed using the following rules: i) the zero vector and any $\mathbf{k}_\phi$ whose first non-zero element is negative are deleted; ii) Every index with a common integer divisor is also deleted. Gallant (1982) shows that $\mathbf{A} = -\sum_{\phi=1}^{\Phi} u_{0\phi}\mathbf{k}_\phi\mathbf{k}'_\phi$. Notice that we can leave $\mathbf{A}$ unrestricted and obtain a different Fourier approximation. We use the same re-scaling as in Gallant (1982) and Feng and Serletis (2009) with the exception that we do not use logs. Finally, the purpose of the re-scaling is that elements of $\mathbf{z}_{it}$ must lie in $[0, 2\pi]$ which can be achieved through a common scale in $\mathbf{z}_{it}$. Finally, we can employ a full third-order approximation:

$$f(\mathbf{z}_{it}) = a_o + \sum_{i=1}^{b} a_i z_{it} + \sum_{i=1}^{b} \sum_{j=1, i \leq j}^{b} b_{ij} z_{it} z_{jt} + \sum_{i=1}^{b} \sum_{j=1}^{b} \sum_{k=1, i \leq j \leq k}^{b} c_{ijk} z_{it} z_{jt} z_{kt}, \tag{35}$$

where b is the number of elements in $\mathbf{z}_{it}$.

We perform posterior analysis with all these specifications, which are flexible enough to cover nearly every empirical case. For any model with parameters $\theta \in \Theta \subseteq \mathbb{R}^{d_\theta}$, including any latent variables in the model, denote the prior by $p(\theta)$, the likelihood by $\mathcal{L}(\theta; \mathcal{Y})$ for data $\mathcal{Y}$ and the posterior by $p(\theta|\mathcal{Y})$. We know

$$p(\theta|\mathcal{Y}) = \frac{\mathcal{L}(\theta; \mathcal{Y}) p(\theta)}{p(\mathcal{Y})}, \tag{36}$$

where the marginal likelihood is

$$p(\mathcal{Y}) = \int_{\Theta} \mathcal{L}(\theta; \mathcal{Y}) p(\theta) d\theta. \tag{37}$$

For two different models, say 1 and 2, we define the Bayes factor in favor of model 1 and against model 2 as:

$$BF_{1:2} = \frac{p_1(\mathcal{Y})}{p_2(\mathcal{Y})}. \tag{38}$$

The Bayes factors, in our application, are reported in Table 2. These results clearly favor the third-order approximation in (35). All of our results will, therefore, be conditional on the selection of this functional form for the productivity equation.

The computation of the marginal likelihood is an involved operation. Here, we compute it using the importance sampling techniques in Perrakis, Ntzoufras and Tsionas (2014).

We also treat firm-specific directions, $\boldsymbol{g}_i$, as latent variables and assume the following prior:

$$\boldsymbol{g}_i \sim \mathscr{N}_{\mathrm{N+M+J+L}} \left(\bar{\boldsymbol{g}}, \boldsymbol{\Sigma}_{\boldsymbol{g}} \right), \tag{39}$$

where $\bar{\boldsymbol{g}}$ denotes the prior mean vector whose elements consist of reasonable prior beliefs (namely, -1 for all inputs and the bad outputs and +1 for the good outputs) and $\boldsymbol{\Sigma}_{\boldsymbol{g}}$ is constructed so as to reflect reasonable deviations from these prior beliefs. Below we find that the posteriors are highly insensitive to variations in these priors.

Having estimated productivity (ω_{it}), we compute PC $= \frac{\partial \omega_{it}}{\partial t}$ and EC $= \frac{\partial u_{it}}{\partial t}$, using equations (30) and (31). Then TC = PC − EC. For more details on computing these quantities, see Grosskopf (2003). We also compute returns to scale (RTS) as the sum of the good output elasticities calculated from the distance function.

3.2.3 Shadow Prices

To incorporate shadow prices as the relevant prices for the firm we write:

$$\frac{p_n^s}{\varrho^s} = \frac{\partial \overrightarrow{D}_{\mathcal{T}}(\boldsymbol{z};\boldsymbol{g})}{\partial x_n},\ n = 1, ..., N, \tag{40}$$

$$\frac{p_m^s}{\varrho^s} = -\frac{\partial \overrightarrow{D}_{\mathcal{T}}(\boldsymbol{z};\boldsymbol{g})}{\partial y_m},\ m = 1, ..., M, \tag{41}$$

$$\frac{p_j^s}{\varrho^s} = \frac{\partial \overrightarrow{D}_{\mathcal{T}}(\boldsymbol{z};\boldsymbol{g})}{\partial \tilde{x}_j},\ j = 1, ..., J, \tag{42}$$

$$\frac{p_l^s}{\varrho^s} = \frac{\partial \overrightarrow{D}_{\mathcal{T}}(\boldsymbol{z};\boldsymbol{g})}{\partial \tilde{y}_l},\ l = 1, ..., L, \tag{43}$$

with $\varrho^s = \sum_{m=1}^{M} p_m^s g_m - \sum_{n=1}^{N} p_n^s g_n - \sum_{j=1}^{J} p_j^s g_j - \sum_{l=1}^{L} p_l^s g_l$, where the superscript s denotes shadow (or perceived) prices to the firm. The shadow prices satisfy the first-order conditions, which are used to complete the system, since we have many endogenous variables but only one DDF equation. Shadow prices are the relevant prices for the shadow-profit-maximizing firm. Actual prices potentially differ from shadow prices by an amount $\xi_q, q = n, m, j, l$:

$$p_n^s = p_n + \xi_n,\ n = 1, ..., N, \tag{44}$$

$$p_m^s = p_m + \xi_m,\ m = 1, ..., M, \tag{45}$$

$$p_j^s = p_j + \xi_j,\ j = 1, ..., J, \tag{46}$$

$$p_l^s = p_l + \xi_l,\ l = 1, ..., L. \tag{47}$$

The variable $\boldsymbol{\xi}_{it}$ is latent, where we assume the following prior:

$$\boldsymbol{\xi}_{it} = [\boldsymbol{\xi}_N', \boldsymbol{\xi}_M', \boldsymbol{\xi}_J', \boldsymbol{\xi}_L']' \sim \mathscr{N}_{N+M+J+L}(\mathbf{0}, \boldsymbol{\Omega}), \tag{48}$$

where $\boldsymbol{\xi}_N = [\xi_{n,it}, n = 1, ..., N]'$, $\boldsymbol{\xi}_M = [\xi_{m,it}, m = 1, ..., M]'$, $\boldsymbol{\xi}_J = [\xi_{j,it}, j = 1, ..., J]'$, $\boldsymbol{\xi}_L = [\xi_{l,it}, l = 1, ..., L]'$, and where $\boldsymbol{\Omega}$ is unknown.

In view of (40)–(43), equations (44)–(47) are rewritten, after introducing stochastic error terms $v_q, q = n, m, j, l$, as:

$$p_n = \varrho^s(\boldsymbol{p}, \boldsymbol{g}, \boldsymbol{\xi}) \cdot \frac{\partial \vec{D}_{\mathcal{T}}(\boldsymbol{z}; \boldsymbol{g})}{\partial x_n} - \xi_n + v_n, n = 1, \ldots, N, \tag{49}$$

$$p_m = -\varrho^s(\boldsymbol{p}, \boldsymbol{g}, \boldsymbol{\xi}) \cdot \frac{\partial \vec{D}_{\mathcal{T}}(\boldsymbol{z}; \boldsymbol{g})}{\partial y_m} - \xi_m + v_m, m = 1, \ldots, M, \tag{50}$$

$$p_j = \varrho^s(\boldsymbol{p}, \boldsymbol{g}, \boldsymbol{\xi}) \cdot \frac{\partial \vec{D}_{\mathcal{T}}(\boldsymbol{z}; \boldsymbol{g})}{\partial \tilde{x}_j} - \xi_j + v_j, j = 1, \ldots, J, \tag{51}$$

$$p_l = \varrho^s(\boldsymbol{p}, \boldsymbol{g}, \boldsymbol{\xi}) \cdot \frac{\partial \vec{D}_{\mathcal{T}}(\boldsymbol{z}; \boldsymbol{g})}{\partial \tilde{y}_l} - \xi_l + v_l, l = 1, \ldots, L, \tag{52}$$

where $\varrho^s(\boldsymbol{p}, \boldsymbol{g}, \boldsymbol{\xi}) = \Big(\varrho(\boldsymbol{p}, \boldsymbol{g}) + \sum_{m=1}^{M} \xi_m g_m - \sum_{n=1}^{N} \xi_n g_n - \sum_{j=1}^{J} \xi_j g_j - \sum_{l=1}^{L} \xi_l g_l\Big)$, the vector $\boldsymbol{\xi} = (\boldsymbol{\xi}_N, \boldsymbol{\xi}_M, \boldsymbol{\xi}_J, \boldsymbol{\xi}_L)$, and the vector $\boldsymbol{p} = (\boldsymbol{p}_N, \boldsymbol{p}_M, \boldsymbol{p}_J, \boldsymbol{p}_L)$. We can normalize the first element of $\boldsymbol{\xi}_N$ to zero, since we can only identify relative price distortions.[12]

Assuming that we have data on prices for all inputs and outputs, estimating all of the normalized price equations in (12)–(15) together with the DDF would generate a singularity in the covariance matrix of the residuals. This occurs since the normalized price equations sum to 1 for any value of $\boldsymbol{g}$ (as is seen from adding (12)-(15) with appropriate changes in sign and comparing to (16)). The choice of which price equation to drop does not impact the results. The technology system is the DDF in (20) substituted into (29) and $N + M + J + L - 1$ of the price equations in (49)–(52), subject to imposition of the restrictions in (27). Relative to one normalized input or output efficiency parameter, we can estimate both input and output price efficiencies by estimating this system, since all the price equations are included.

3.2.4 Completing the System using Reduced-Form Equations

Assume now a worse-case scenario where we lack prices and alternative valid instruments for the endogenous bads, $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{y}}$, so that we are unable to identify the DDF using equations like (51) and (52). Assume further that we are unable to generate latent prices for $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{y}}$. We can still identify the technology DDF. First we create a system of equations consisting of (20) substituted into (29) and $N + M - 1$ of the price equations in (49) and (50), subject to the restrictions in (27). Then we complete this system by including reduced-form equations for $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{y}}$ following standard LIML practices as in

[12] This occurs since PM means that the firm chooses absolute levels of outputs given CM, which requires that the firm equate ratios of input prices to ratios of marginal products. Estimation of the extent to which the latter has been achieved requires one normalization.

Pagan (1979). The input and output DDF systems would be constructed analogously. The explanatory variables for these equations are prices of the good inputs and good outputs, firm dummies, and time dummies. See Appendix B for further details.

3.2.5 Completing the System using Latent-Price Equations

Frequently, actual prices for inputs and outputs are missing. With rare exceptions, prices for good inputs are confidential for privately-owned, unregulated firms. While data on prices and quantities of good inputs and outputs are typically reported by regulated electric utilities, the federal government currently allows many utilities to redact data on wages. Publicly-available data sources intermittently report prices from thin tradable permit markets for the regulated bad outputs that we examine, sulfur dioxide (SO_2) and nitrogen oxide (NO_x). No reliable prices exist for the unregulated pollutant that we model, carbon dioxide (CO_2). Further, prices of bad inputs are missing for all firms. In theory one could estimate prices for sulfur, a bad input, which is chemically bound to coal and oil, which are good fuel inputs. However, this would require the use of hedonic methods, which are infeasible due to data confidentiality.[13]

In place of missing actual prices, we can sample prices from their conditional posteriors generated by their FOCs to generate latent prices $\mathbf{p}_j^*$ and $\mathbf{p}_l^*$ for $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{y}}$ to replace the missing $\mathbf{p}_j$ and $\mathbf{p}_l$ in (51) and (52), respectively. Then we can estimate a complete (fully-identified) system comprised of (20) substituted into (29) and (49)–(52), subject to (27).[14]

Assuming that input prices are unobserved and, therefore, generated as latent variables requires stochastic assumptions. Of course, these assumptions have to be consistent with what is known about the sector under study and the nature of sectoral input prices.

Let $\mathbf{p}^* = [p_l^*, l = 1, \ldots, L; p_j^*, j = 1, \ldots, J]$. Assume there is a $Q \times 1$ vector of predetermined variables $\mathbf{f}_{it}$, so that $\mathbb{E}(p_{it,h}^*|\mathbf{f}_{it}, \zeta_h) = \mathbf{f}_{it}'\zeta_h$ for $h = 1, \ldots, J + L$ where

[13] In a competitive market, the delivered price of coal is a function of the Btu/ton, the percent sulfur/ton, and transportation charges. Given this information, a hedonic regression would yield the implicit price of the percent sulfur/ton. Unfortunately, the DOE publishes data on all these variables except for transportation charges, which are confidential. A way around this would be to obtain data on the mine-mouth price/ton, which could be regressed on Btu/ton and percent sulfur/ton. Again unfortunately, complete mine-mouth data on price/ton linked to a specific mine and utility is confidential.

[14] One can apply this approach if prices are also missing for elements of $\mathbf{x}$ and $\mathbf{y}$ by simply generalizing the following procedure.

$\zeta_h, h = 1, \ldots, J+L$ are $Q \times 1$ parameter vectors. Define $\mathbf{F}_{it} = I_{J+L} \otimes \mathbf{f}_{it}$ so that

$$\boldsymbol{p}^*_{it} = \mathbf{F}_{it}\zeta + \epsilon_{it},\ \epsilon_{it} \sim \mathcal{N}_{J+L}(0, \Sigma_{p^*}). \tag{53}$$

Thus, the predetermined variables are in $\mathbf{F}_{it}$.

This can be written in the standard multivariate regression form as $\boldsymbol{p}^* = \mathbf{F}\zeta + \epsilon$. For a particular observation we assume

$$\boldsymbol{p}^*_{it}|\mathbf{F}_{it}, \zeta \sim \mathcal{N}_{J+L}(\mathbf{F}_{it}\zeta, \Sigma_{p^*}), \tag{54}$$

where $\zeta = [\zeta'_1, \ldots, \zeta'_{J+L}]'$ is $Q' \times 1$ where $Q' = (J+L)Q$. As we lack specific prior information, we assume:

$$p(\zeta, \Sigma_{p^*}) \propto f_{N,J+L}(\zeta; \bar{\zeta}, \bar{V}_\zeta) \cdot |\Sigma_{p^*}|^{-(J+L+1)/2}, \tag{55}$$

where $f_{N,J+L}(\zeta; \bar{\zeta}, \bar{V}_\zeta)$ denotes a $(J+L)-$dimensional normal density with mean vector $\bar{\zeta} = \mathbf{0}_{J+L}$ and covariance $\bar{V}_\zeta = 10^4 I_{J+L}$. Relative to the previous approach, we now have to draw from three additional conditional posterior distributions, viz.:

$$\begin{aligned} &\boldsymbol{p}^*|\zeta, \Sigma_{p^*}, \ldots, \\ &\zeta|\boldsymbol{p}^*, \Sigma_{p^*}, \ldots, \\ &\Sigma_{p^*}|\boldsymbol{p}^*, \zeta, \ldots. \end{aligned} \tag{56}$$

All three additional conditional posterior distributions are in standard form. Standard results yield the following:

$$\zeta|\cdot \sim \mathcal{N}_{Q'}(\hat{\zeta}, \hat{V}), \tag{57}$$

where $\hat{\zeta} = \left(\mathbf{F}'(I \otimes \Sigma^{-1}_{p^*})^{-1}\mathbf{F}\right)^{-1} \mathbf{F}'(I \otimes \Sigma^{-1}_{p^*})^{-1}\boldsymbol{p}^*$ and $\hat{V} = \left(\mathbf{F}'(I \otimes \Sigma^{-1}_{p^*})^{-1}\mathbf{F}\right)^{-1}$. For the covariance matrix we have:

$$p(\Sigma_{p^*}|\cdot) \propto |\Sigma_{p^*}|^{-(nT+J+L+1)/2} \exp\left(-\mathrm{tr}\bar{A}\Sigma^{-1}_{p^*}\right), \tag{58}$$

where $\bar{A} = (\boldsymbol{p}^* - \mathbf{F}\zeta)'(\boldsymbol{p}^* - \mathbf{F}\zeta)$.

Assuming the covariance matrix of errors in (51) and (52) is $\tilde{\Sigma}$ we obtain that

$$\boldsymbol{p}^*_{it}|\cdot \sim \mathcal{N}_{J+L}(\bar{p}_{it}, \bar{V}_{p^*}), \tag{59}$$

where $\bar{p}_{it} = (\Sigma^{-1}_{p^*} + \tilde{\Sigma}^{-1})^{-1}(\Sigma^{-1}_{p^*}\mathbf{F}_{it}\zeta + \tilde{\Sigma}^{-1}\tilde{\mathbf{G}}_{it})$ and $\bar{V}_{p^*} = (\Sigma^{-1}_{p^*} + \tilde{\Sigma}^{-1})^{-1}$, where $\tilde{\mathbf{G}}_{it}$ denotes the RHS of (51) and (52).[15]

[15] Coupled with (27), the price equations in (49)–(52) satisfy the order condition for identification. Of course this discussion is confined to a frequentist view of identification. In Bayesian models, even unidentified parameters can be identified with proper priors. In Table 3 and figures 10 and 11 we provide information on the sensitivity of posterior densities to the choice of priors.

3.3 Measurement of Allocative Inefficiency and its Resource Implications

Using (49)–(52) we can measure allocative inefficiency (AI) as the sum of the error in each input price equation times its corresponding quantity plus the sum of the error in each output price times its corresponding quantity all divided by $\varrho(\boldsymbol{p}, \boldsymbol{g})$ (defined in (16)):

$$AI_{it} = \left(\sum_{n=1}^{N} \xi_{n,it} x_{n,it} + \sum_{m=1}^{M} \xi_{m,it} y_{m,it} + \sum_{j=1}^{J} \xi_{j,it} \tilde{x}_{j,it} + \sum_{l=1}^{L} \xi_{l,it} \tilde{y}_{l,it} \right) / \varrho(\boldsymbol{p}, \boldsymbol{g}). \tag{60}$$

To reiterate, this measure depends on the errors from (49)–(52) weighted by corresponding quantities. Since we interpret ϱ as the change in profits due to an incremental improvement in the production technology, we can decompose ϱ into two parts: that due to reducing price inefficiency and that due to reducing technical inefficiency. The numerator of AI_{it} is the former component.

To account for parameter uncertainty, AI_{it} is averaged across Markov Chain Monte Carlo (MCMC) draws in standard Rao-Blackwell fashion. For each MCMC draw, AI_{it} in (60) can be transformed as $\widetilde{AI}_{it} := AI_{it} - \min\{AI_{it}\}$. Allocative efficiency (AE) is then computed as

$$AE_{it} = 100 - \widetilde{AI}_{it}. \tag{61}$$

Since we make this computation for each MCMC draw, the probability of any one firm being fully efficient is very small. Our procedure allows us to resolve the standard problem of relative efficiency estimation by using our variant of the corrected ordinary least-squares technique in a Bayesian framework.

We also compute the estimated percent change in input usage for inputs and outputs due to the firm producing at profit-maximizing levels based on market prices rather than profit-maximizing levels based on shadow prices. For each of equations (49)–(52) we solve the linear system for each draw of parameters and latent variables, which include the estimated value of $\boldsymbol{\xi}$. A solution $\mathbf{z}_{*,o}^{s}$ is obtained for iteration $s, s = 1, \ldots, S$, for all observations $o = 1, ..., \mathcal{N}T$. This measures the optimal level of inputs subject to market prices, since the solved equations are in terms of these prices. The amounts $\mathbf{z}$ are the current optimal amounts of input usage subject to shadow prices. Then we take the Monte Carlo average $\mathbf{z}_{*,o} = S^{-1} \sum_{s=1}^{S} \mathbf{z}_{*,o}^{s}$ to account for parameter uncertainty. Finally, we generate the sample distribution of changes $\mathbf{c}_o$, where $\mathbf{c}_o = \frac{\mathbf{z}_{*,o} - \mathbf{z}_o}{\mathbf{z}_o}$.

3.4 Posterior Predictive Measure of Cost-Minimization versus Profit Maximization

As indicated above, one can obtain the CM model by estimating the DDF with only the good and bad input price equations. This entails estimation of (20) substituted into (29), (49), and (51), where (27) is imposed during our MCMC iterations.

We wish to be able to compare the CM model to the PM model in terms of their predictive abilities. Suppose all T observations (say $y_i \in \Re^T$) of a certain utility $i \in \{1, \ldots, \mathcal{N}\}$ are omitted from the dataset Y so the new dataset is Y_{-i}. The new posterior is $p(\theta_{(i)}|Y_{-i})$ and the posterior predictive distribution for the ith utility is:

$$p(\tilde{y}_i|Y_{-i}) = \int p(\tilde{y}_i|\theta_{(i)}, Y_{-i})p(\theta_{(i)}|Y_{-i})d\theta_{(i)}. \tag{62}$$

We are interested in the posterior distribution of the absolute prediction error:

$$APE = \int_{\Re^T} |y_i - \tilde{y}_i| p(\tilde{y}_i|Y_{-i})d\tilde{y}_i. \tag{63}$$

This can be computed with a simple Monte Carlo simulation:

A. Obtain $\theta_i^s, s = 1, ..., S$ using MCMC.

B. Draw $\tilde{y}_i^s, s = 1, ..., S$ from the likelihood, viz., the distribution of y_i given $\theta_{(i)}^s$.

C. Compute $APE \simeq S^{-1} \sum_{s=1}^{S} |y_i - \tilde{y}_i^s|$.

This can be obtained for both the CM and PM models. We use as a predictive measure the ratio:

$$R = \frac{APE_{PM}}{APE_{CM}}. \tag{64}$$

In Fig. 13 we present predictive R for our sample.

Apparently, the vector y_i must include only the endogenous variables that are common in the two models. A ratio that has most posterior probability mass $R < 1$ would indicate that the PM model does a better job in terms of posterior prediction. In step 1, the MCMC is implemented using the same size of burn-in and subsequent draws to convergence as in the original MCMC simulations using the PM and CM model.

3.5 Comparison of PM Models with and without Latent Prices

We compare three shadow-PM models that estimate the DDF together with ancillary equations. All models include the quadratic DDF (20) substituted into (29). We also include in all models two good input price equations containing price inefficiency terms from (49) for capital and energy (dropping that for labor), and two good output price equations from (50) for residential and industrial/commercial electricity production. With Model I, we ignore endogeneity and simply estimate this system without instruments. With Model II, we identify the DDF using reduced-form equations for the bad input and the three bad outputs. In Appendix B we explain in more detail the reduced-form equations, whose explanatory variables are always the prices of the good inputs and good outputs, firm dummies, and time dummies.

For Model III we identify the DDF by utilizing the specification for Model II, but replacing the reduced-form equations with latent price equations from (52) for the two bad outputs, SO_2 and NO_x, which are regulated. To determine which variables should be included in (53), we argue that in equilibrium, the price of each regulated bad output for the firm should equal its marginal cost of control (which should also equal the price of an emissions permit if the emission constraint is binding). The marginal cost of control of each bad output is a function of exogenous prices of inputs used to control that bad output: the prices of capital, labor, and energy.[16]

Thus, for Model III, we estimate latent prices for SO_2 and NO_x as functions of firm dummies, time dummies, vintage, and the prices of the good inputs, which comprise $\mathbf{F}_{it}$ in (53). We do not include a first-order price equation for sulfur, since it is purchased jointly with the good input, energy, and we lack the data to compute the negative hedonic price of sulfur as discussed above. Also, we omit the price equation for CO_2, since it is an unregulated pollutant with zero price to the firm. We complete this system by specifying reduced-form equations for the endogenous variables, the quantities of sulfur and CO_2. The translation property restrictions for a technology DDF from (27) are substituted into all equations except for the reduced-form ones. Estimating the DDF in conjunction with first-order price equations in terms of shadow prices means that the directions are estimated subject to PM conditions (satisfied with near equality). In this sense, the estimated directions are optimal. We summarize Bayes factor comparisons of our models in Fig. 2 discussed below.

We follow the practice of computing and reporting predictive Bayes factors instead of reporting a single Bayes factor for the entire sample since the possibility always exists

[16] Our data contains an overall price for each good input, but does not provide separate prices for the portion used to reduce bad outputs and increase good outputs.

that some observations are particularly influential in driving the Bayes factor in favor of a particular model. For our data, this does not appear to be the case.

4 Econometric Implementation

4.1 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. This data set is 80% longer than that used by AT, which runs from 1988-1997, thereby omitting a period when many states restructured (deregulated) the generation and delivery of electricity. Approximately 11% of our observations are for periods of restructuring.[17] A number of firms either merged or sold their assets and dropped out of the sample from 1998 onward. A balanced panel would have yielded 1386 observations.

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. We include a full set of 77 firm-specific dummies and omit the intercept in the DDF (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) and three bad outputs, SO_2, CO_2, and NO_x. From *Federal Energy Regulatory Commission (FERC) Form 1* (1988-2005) we obtain the quantity of energy in total Btu of fuel consumed and the quantity of labor as the number of full-time plus one-half the number of part time employees. In this form, utilities report total capital expenses as the dollars of interest plus depreciation paid by each utility for the sum of production capital and pollution-control capital. From this same form, we decompose total generation into residential and industrial/commercial generation by multiplying total steam output by the percent of sales revenue in each category.

[17] For a summary of the goals from restructuring the electricity utility industry see Borenstein and Bushnell (2015).

We also calculate 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, taken from *Department of Energy Information Administration (EIA) Form 767 Boiler Files.* The price of labor 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 taken from *FERC Form 1* (1988-2005). The price of capital is 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. These data were taken from *FERC Form 1* (1988-2005) and *Moody's Public Utility Manual* (1988-2005). The prices of residential and industrial/commercial production are derived as total revenues in each category divided by total sales in that category, where the data are taken from *FERC Form 1* (1988-2005).

Data are available on the quantities, but not prices, for the bad input (sulfur) consumed and all bad outputs generated by the firm. These data are obtained from the *EIA Form 767 Boiler Files.*

We compute τ (vintage) using the weighted-average age in years for the firm's capital, where weights are computed using the firm's kilowatt-hour (kWh) output, taken from *FERC Form 1* (1988-2005)). This variable and the time dummies are found 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.

Consumption of total kWh by industrial/commercial customers (66%) was considerably larger than that of residential users (34%) over the sample period. Over the years 1988-1998, total generation remained relatively constant for our sample firms. However, in 1999 total kWh production began a steady decline through the year 2005. For our sample, SO_2/kWh has fallen by about 35% over the sample period in response to the increasingly tight emission restrictions under the SO_2 cap-and-trade system. While NO_x/kWh has fallen slightly, CO_2/kWh has risen significantly, since this pollutant is unregulated. All continuous variables are standardized to eliminate variation due to different units of measure when computing directions.

4.2 Satisfying Properties D1-D8

As indicated above, we satisfy the translation property of the DDF, D1, by imposing on each estimated model the restrictions in (27). The restriction D1 implies D2 for our estimated system. To see this, first impose the restrictions in (27) on (25) (with zero on the left-hand-side in order to guarantee frontier efficiency), using $\vartheta_o = -1, \vartheta = 0$, and scale $\boldsymbol{g}$ by λ. We obtain

$$0 = \overrightarrow{D}_{\mathcal{T}}(\mathbf{z} + \lambda\tilde{\alpha}\boldsymbol{g}; \boldsymbol{g}) \quad = \quad \overrightarrow{D}_{\mathcal{T}}(\mathbf{z}; \boldsymbol{g}) - \lambda\tilde{\alpha}, \tag{65}$$

where $\tilde{\alpha}$ is the new estimated distance (which obtains after scaling $\boldsymbol{g}$ by λ). This implies that

$$\tilde{\alpha} = \lambda^{-1}\overrightarrow{D}_{\mathcal{T}}(\mathbf{z}; \boldsymbol{g}), \tag{66}$$

which says that the new estimated distance, $\tilde{\alpha}$, equals λ^{-1} times the original distance, which is a function of the original $\boldsymbol{g}$. A similar demonstration can be made for the input and output models. We randomly test for concavity, D3, and find that it is satisfied 99% of the time. Non-negativity, D4, is imposed after estimation for all observations via a normalization of the fitted DDF. The monotonicity properties–D5-S, D6-S, D7-W, and D8-W–are satisfied for nearly all of the data via the MCMC estimation process, as explained in the following subsection.

4.3 Statistical Inference in the PM and CM Systems

Our implementation of MCMC relies on a burn-in or transient phase whose length is determined by using Geweke's diagnostic (1992). In preliminary experiments with various priors, the length of the transient phase ranged from 250,000 to 500,000 iterations. For the baseline prior we used 500,000 burn-in draws followed by another million draws which we use to compute marginal posterior densities and statistics for the functions of interest.

We include a restructuring dummy and also consider the interactions of the restructuring dummy with all first-order terms. Although the restructuring dummy itself is significant with a p-value of .0031, a traditional F-test of the interactions as a whole has a p-value of 0.230. This is really a Bayesian F-test which averages across all MCMC draws.

Monotonicity constraints are often violated in empirical applications. In this paper, we impose the monotonicity constraints at the means of the variables (which are normalized to zero) and also at a number of points whose distance from the mean is r. Since the data are normalized to have unit standard deviation, values of r up to 3 are considered.

Pertaining to all monotonicity properties, the number of violations for the baseline prior is very similar to that for the other priors. Without any restrictions, we have 68 violations. When the restrictions are imposed at the mean we have 31. When the restrictions are imposed at the mean plus a point which is $r = 0.5$ units away from the mean we have 11 violations. Imposing the restrictions with $r = 2$ we have zero violations at the posterior mean of the parameters and a maximum of 2 violations across all MCMC draws. We employ this value for r in the results reported below. The imposition of monotonicity constraints is done using standard rejection techniques.

Our figures and tables pertain only to the shadow-PM system, with the exception of Fig. 13, which indicates that PM is superior to CM. Table 1 summarizes our identification strategies for Models I-III. In Appendix B we explain our use of a Fourier expansion to generate reduced-form expressions used in Models II and III. Based on Bayes factors, increasing the number of terms beyond three provides little improvement. Hence, we employ three terms in our expansion. Figs. 2a and 2b summarize the log of the predictive Bayes Factors against Models I and II and in favor of Model III. Panel 2a (2b) compares the models by omitting utilities (time periods). Clearly Model III is superior based on Bayes factors.

Table 2 reports Bayes factors for different specifications of the productivity equation (30) relative to the linear model, which we employ as model 2 in (38). Throughout the paper, we employ the third-order specification, since it clearly dominates with a Bayes factor of 12.55.[18]

Figs. 3a and 3b present posteriors of the optimal directions of inputs and outputs for Model III. The mean directions for capital, labor, and residential generation are negative, with posterior means of about -.4, -.65, and -.7, respectively. The mean di-

[18] Doraszelski, U. and J. Jaumandreu (2013) avoid assuming that all prices are exogenous by omitting subsets of moments involving either lagged wage or lagged price of materials. We have not dropped subsets of moments involving prices since we are assuming that all inputs and outputs are potentially endogenous and we need all prices as instruments. To consider whether lagged prices are better instruments than current period prices, we employ one-period lagged prices for Model III in place of current-year prices. The third-order specification of the productivity equation with current year prices is still superior, in this case with a Bayes factor of 15.66.

rection for energy is positive, with a posterior mean of about .4. Further, the mean direction for the other good and bad outputs are positive. The largest of these is for industrial/commercial output with a mean of about .35. The positive mean direction for energy and the negative mean direction for residential generation are the opposite signs of what is normally assumed. However, since energy is an input substitute and industrial/commercial generation is an output substitute, their signs are reasonable. In addition, the positive directions for industrial/commercial output, NO_x, and SO_2 emissions are consistent with the assumption that the bads are weakly disposable relative to industrial/commercial output. Model misspecification would occur if one arbitrarily assigns directions for bad outputs and all inputs of -1 and directions for good outputs of +1 or requires that industrial/commercial output and the bad outputs have directions with opposite signs.

Given estimated optimal-PM directions for measuring distances from the frontier, in Fig. 4 we present the posterior density of TE for models I-III. Models I and II indicate mean TE of about .75 and .65, respectively. With Model III, TE is about .85.

Fig. 5 provides for Model III the posterior percentage price distortions (relative to actual prices) for good inputs and good outputs. The posterior means are small and negative for all good inputs. From equation (49) the negative price distortions for inputs indicate that their shadow prices are less than their actual prices. Specifically, the average shadow price of capital is less than its market price due to super-normal rates of return allowed in many rate-of-return regulated regimes, which dominate our sample. This result is consistent with those of other studies of electric utilities referenced above. From equation (50), somewhat larger positive mean price distortions for good outputs imply that their shadow prices are slightly greater than their actual prices.

Fig. 6 shows for Model III the posterior percent price distortions (relative to latent prices) for bad outputs, computed using (52). Mean price distortions are small and positive for both bad outputs, indicating that shadow prices of the bad outputs exceed their market prices.

In Fig. 7, for Models I-III we provide the posterior densities of AE computed from equation (61) for all inputs and outputs. Mean AE for Models I-III is about .85, .66, and .64, respectively.

In Fig. 8 we provide for Model III the posteriors of the percent changes in the usage of inputs and outputs if firms were required to produce at profit-maximizing levels based on actual (market) prices rather than profit-maximizing levels based on shadow prices. We use the methodology described in section 3.3 above. The computed percent change for a given input or output is a function of marginal rates of transformation and price changes of all other inputs and outputs, through the solution of a system of equations. Thus, the percent changes in usage are not linearly related to individual price distortions. The mean posterior changes in usage are -.04 and -.07 for capital and labor, respectively, and .05 for energy. The mean posterior changes for good and bad outputs range from .02 to .04. All of these distortions are relatively small.

Fig. 9 provides posterior densities for RTS, PC, TC, and EC for Model III. The posterior mean of RTS is about .94, indicating slightly decreasing returns to scale. The posterior mean of EC is slightly negative and its posterior distribution is somewhat less disperse than are those of PC and TC. The posterior mean of TC is slightly greater than .01 and the posterior mean of PC is slightly less than .01.

Table 3 indicates the range in 10,000 priors used for posterior sensitivity analysis. Fig. 10 reports the resulting posteriors for inputs and outputs for Model III. We focus on changes in the posteriors of inputs and outputs relative to the baseline prior. To minimize computational costs of this posterior comparison for each of the 10,000 different priors, we use Sampling-Importance-Resampling (SIR) following Smith and Gelfand (1992) and Rubin (1987). Given a set of posterior draws $\left\{\boldsymbol{\beta}^{(s)}, s=1,...,S\right\}$ for a model with a given baseline prior, say $p_o\left(\boldsymbol{\beta}\right)$, approximate draws from the same model with an alternative prior $p\left(\boldsymbol{\beta}\right)$ can be obtained using the SIR algorithm. This attaches weights, $\Upsilon_s = \frac{p\left(\boldsymbol{\beta}^{(s)}\right)}{p_o\left(\boldsymbol{\beta}^{(s)}\right)}$, to the original draws and resampling is used with these normalized weights avoiding the reuse of MCMC[19]. All sensitivities are quite small.

Fig. 11 reports the sensitivity of changes in posterior means for structural parameters of Models I-III to 10,000 different priors. Since the use of SIR does not require new MCMC computations, it is particularly suited to large-scale prior sensitivity analysis as in our case. The 10,000 different priors are generated from the baseline prior of each parameter or block using the hyperparameters of these priors. If the vector of hyperparameters is collectively denoted by $\delta \in \mathbb{R}^{\Delta}$ new priors are generated using $\delta^* = \delta + \kappa$ where κ is uniformly distributed in $[-B_1, B_2]^{\Delta}$.

We set $B_1 = B_2 = 10$ for hyperparameters that can be defined over the real line and $B_1 = 0.001$, $B_2 = 10$ for positive hyperparameters. Performing the same sensitivity

[19]The size of the resample is set to 20% of the number of available MCMC draws.

analysis experiment when only the priors for the structural parameters are allowed to change, the changes in posterior means are modest, suggesting that posterior MCMC is quite robust.

After estimating the shadow PM model, we compute for Model III the posterior means of relative shadow prices over time for bad outputs from equation (52). As reported in Fig. 12, these prices all decline over time, consistent with historically declining costs of pollution control. The relative prices of SO_2 and NO_x are consistent with estimates of control costs from the Integrated Environmental Control Model (Rubin, 2009).

In order to compare the predictive accuracy of the shadow-CM and shadow-PM models, we compute the marginal posterior of the predictive measures of R. For five randomly chosen utilities, provided in Fig. 13, R values range from zero to .5. The mean values range from approximately .01 to about .1, indicating that the shadow-PM model is strongly preferred to the shadow-CM model.

Finally, in Figs. 14-15 and Table 4 we report for Model III the partial elasticities of ω in (30) with respect to $\omega_{t-1}, \mathbf{p}_{\mathrm{x}}, \mathbf{p}_{\mathrm{y}}, \pi, \tau$ and t. When implementing (32)–(35) we do not impose monotonicity of profits with respect to ω. However, we find that the required monotonicity is satisfied for 99% of our observations.

Lagged ω is the most important variable affecting ω with a positive elasticity of .43. Reducing p_E is the second most important variable with an elasticity of -.41. Reductions in p_K and p_L are more important than an increase in the prices of residential and industrial production. Profits have a very small but positive effect on productivity. All elasticities except for those of t and τ are significant at the .05 level using a two-tailed test, indicating little affect of time itself or the aging of capital stock.

5 Conclusions

Using a Bayesian appproach, our contributions to the productivity literature are fourfold. First, we estimate unique optimal-PM directions for each good input, each good output, and all regulated bad outputs. Second, we allow the potential endogeneity of all inputs and outputs. This entails identifying the DDF by assuming a data generating process for the latent prices as instruments to replace missing prices for the regulated bad outputs. Then we estimate the DDF jointly with the first-order price equations derived from PM for good inputs, good outputs, and two controlled pollutants. The bad input, sulfur, and the uncontrolled pollutant, CO_2, are identified using reduced-form equations. We also estimate the corresponding CM model. Third, we avoid the typical assumption that firms respond to actual prices by estimating firm-specific shadow prices. Fourth, we generalize and provide an alternative to the approaches of OP (1996) and LP (2003). We accomplish this by treating all input and output quantities as potentially endogenous and deriving productivity as a function of lagged productivity, profits, vintage, time, and the prices of good inputs and outputs. From this function, we compute TC, EC, PC, and partial elasticities.

Using MCMC methods, we generate posterior 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. Using Bayesian criteria, the shadow PM model is superior to the shadow CM model in terms of predictive ability. Optimal directions subject to shadow PM differ from their typically assumed values (-1 for all bad outputs and all inputs, and +1 for good outputs). Estimated price distortions for the good inputs indicate that efficient levels of usage are slightly lower for capital and labor but are moderately higher for energy. Mean PC is slightly less than .01. The major factors augmenting productivity are an increase in lagged productivity and a reduction in energy prices. Changes in posterior means of the structural parameters and the directions are highly insensitive to a wide range of different priors.

References

Ackerberg, D. A., Caves, K., Frazer, G., 2015, Identification properties of recent production function estimators Econometrica 83, 2411-51.

Atkinson, S.E., Halvorsen, R., 1976, Interfuel Substitution in Steam-Electric Power Generation. Journal of Political Economy 84, 959-78.

Atkinson, S.E., Halvorsen, R., 1980, A test of relative and absolute price efficiency in regulated utilities, Review of Economics and Statistics 62 (1), 81-88.

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

Borenstein, S., Bushnell, J., 2015, The U.S. electricity industry after 20 years of restructuring. Annual Review of Economics 7, 437-463.

Caves, D.W., Christensen, L. R., Diewert, W. E., 1982a, The economic theory of index numbers and the mesurement of input, output, and productivity. Econometrica 50, 1393-1414.

Caves, D.W., Christensen, L. R., Diewert, W. E., 1982b, Multilateral comparisons of output, input, and productivity using superlative indexes. The Economic Journal 92, 73-86.

Chambers, R. G., 1998, Input and Output Indicators. In: Färe, R., Grosskopf, S., Russell, R. (Eds.), Index Numbers: Essays in Honor of Sten Malmquist Boston: Kluwer Academic Publishers , 241-271.

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

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

Cowing, T.G., 1978, The effectiveness of rate-of-return regulation: an empirical test using profit functions. In: M. A. Fuss, McFadden, D.L. (Eds.), Production Economics: A Dual Approach to Theory and Applications, Amsterdam: North Holland.

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

Doraszelski, U., Jaumandreu, J., 2013, R & D and Productivity: Estimating Endogenous Productivity. Review of Economic Studies 80, 1338-1383.

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

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

Färe, R., Lundberg, A., 2005, Parametrizing the Shortage Function, working paper, Department of Economics. & Agricultural and Resource Economics, Oregon State University.

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

Feng, G., Serletis, A., 2009, Efficiency and productivity of the US banking industry, 1998–2005: evidence from the Fourier cost function satisfying global regularity conditions. Journal of Applied Econometrics, 24, 105-138.

Fowlie, M., 2010, Emissions Trading, Electricity Restructuring, and Investment in Pollution Abatement. American Economic Review 100, 837-69.

Gallant A. R., 1982, Unbiased determination of production technology. Journal of Econometrics 20(2): 285–323.

Geweke, J., 1992, Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. In: Bayesian Statistics 4 Bernardo, J.M., Berger, J., Dawid, A.P., Smith, A.F.M.) (Eds.). Oxford: Oxford University Press, 169-193.

Griliches, Z., Mairesse, J., 1998, Production functions: the search for identification. In: Strom S., (Ed.) Econometrics and Economic Theory in the 20th century. Strøm, S. ed., Cambridge University Press, pp. 169-203.

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

Hampf, B., Krüger, J., 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., Primont, D. 2007, Derivative Properties of Directional Technology Distance Functions. In: Färe, R., Grosskopf, S., Primont, D. (Eds.). Aggregation, Efficiency, and Measurement, Springer.

Kumbhakar, S.C., Lovell, C. A. K., 2000, Stochastic Frontier Analysis, Cambridge.

Levinsohn, J., Petrin, A., 2003, Estimating production functions using inputs to control for unobservables, Review of Economic Studies 70, 317-342.

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

Olley, G. S., Pakes, A., 1996, The dynamics of productivity in the telecommunications equipment industry. Econometrica 64, 1263-1297.

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

Perrakis, K., Ntzoufras, I., Tsionas, M. G., 2014, On the use of marginal posteriors in marginal likelihood estimation via importance sampling. Computational Statistics & Data Analysis 77, 54-69.

Pope, R. D., Just, R. E., 1996, Empirical implementation of ex ante cost functions. Journal of Econometrics 72, 231-249.

Rubin, E. S., 2009, Integrated Environmental Control Module (IECM), version 6.2.4, (developed under contracts from DOE), Carnegie Mellon University.

Rubin, D.B., 1987, The calculation of posterior distributions by data augmentation. Journal of the American Statistical Association 82, 543-546.

Smith, A. F. M., Gelfand, A. E., 1992, Bayesian statistics without tears: a sampling–resampling perspective. The American Statistician 46, 84-88.

Wang, H-J, Schmidt, P., 2002, One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels. Journal of Productivity Analysis 18, 129-144.

Wooldridge, J. M., 2009, On estimating firm-level production functions using proxy variables to control for unobservables. Economic Letters 104, 112-14.

Yaisawarng, S., Klein, J. D., 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.

Zofio, J. L., Pastor, J. T., Aparicio, J., 2013, The Directional Profit Efficiency Measure: On Why Profit Inefficiency is Either Technical or Allocative. Journal of Productivity Analysis 40, 257-266.

6 Tables and Figures

Table 1: Identification Strategies for Models I-III

Model	Strategy	Variables with Identifying Equations
Model I	Ignore Endogeneity	none
Model II	Add Reduced-Form Equations	$\tilde{x}_S$ and $\tilde{\mathbf{y}}$
Model III	Add Reduced-Form Equations Add Latent-Price Equations	$\tilde{x}_{\mathrm{S}}$, $\tilde{y}_{\mathrm{CO_2}}$ $\tilde{p}_{\mathrm{SO_2}}, \tilde{p}_{\mathrm{NO_x}}$

Table 2. Bayes factors for different specifications of productivity equation

	BF
linear	1.000
second-order	5.301
third-order	12.545
fourth-order	2.271
NN, $G = 1$	6.412
NN, $G = 2$	5.812
NN, $G = 3$	4.023
NN, $G = 4$	1.167
Fourier, unrestricted $\mathbf{A}$	2.519

Notes: The linear model is obtained from (33) by setting $\mathbf{A} = \mathbf{0}$, "NN" stands for "neural network" and the specification is given by (32). The Fourier approximation is as in (34). The model "Fourier, unrestricted $\mathbf{A}$" corresponds to (34) with an unrestricted $\mathbf{A}$ matrix.

Table 3. Baseline priors and range of variation for parameters to generate the 10,000 priors for posterior sensitivity analysis

Parameter	Equation	Value	Range of variation
$\boldsymbol{\gamma}\|\boldsymbol{g}$	Appendix (C.1.5)	Semi-informative form	fixed
c	(27)	10^{-4}	10^{-3} to 10^{-8}
$\bar{\boldsymbol{g}}$	(39)	-1 or +1	fixed
$\boldsymbol{\Sigma_g}$	(39)	Wishart$(\underline{\nu}, \underline{\boldsymbol{A}})$	See note below
$\boldsymbol{\Delta}$	Appendix (B.1)	Wishart$(\underline{\nu}, \underline{\boldsymbol{A}})$	See note below
$\boldsymbol{\pi}$	Appendix (B.1)	10^3 varies from 10 to 10^5	10^{-2} varies from 10^{-4} to 1
σ_u	Appendix (C.2.5.1)	$\frac{\underline{Q}}{\sigma_u^2} \sim \chi^2_{\underline{\nu}}$, $\underline{Q} = 0.1, \underline{\nu} = 1$	$Q \in \left[10^{-3}, 100\right]$, $\nu \in [0.01, 20]$
F	Appendix (B.2)	5	Chosen initially using BIC

Note: Wishart priors are of the form $p(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{(\underline{\nu}-m-1)/2} \exp\left(tr\underline{\boldsymbol{A}}^{-1}\boldsymbol{\Sigma}\right)$ where the dimensionality of the matrix is $m \times m$ generically, and $\underline{\nu}, \underline{\boldsymbol{A}}$ are prior parameters. We set the parameter $\underline{\nu}$ equal to 0.1 times the sample size and $\underline{\boldsymbol{A}}^{-1} = 0.001\boldsymbol{I}$.

Table 4. Partial Elasticity of ω_{it} with respect to $\mathbf{p}_x, \mathbf{p}_y, \pi, \tau, t$

$\omega_{i,t-1}$	0.432 (0.081)
p_E	-0.414 (0.122)
p_L	-0.235 (0.092)
p_K	-0.167 (0.072)
p_{RES}	0.045 (0.023)
p_{IND}	0.071 (0.022)
π_{it}	0.044 (0.011)
t	0.0014 (0.0013)
τ_{it}	-0.157 (0.032)

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

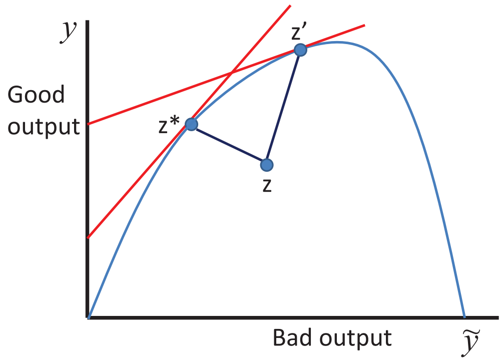

Fig. 2: Log of Predictive Bayes Factors

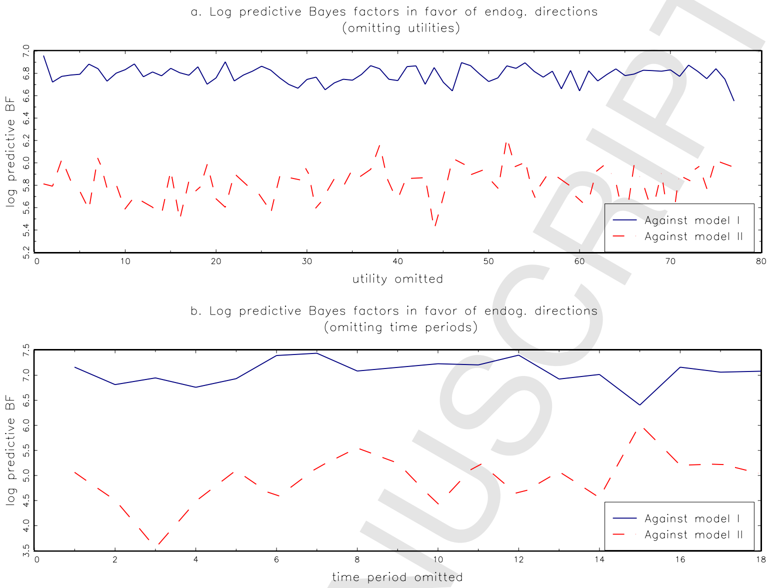

Fig. 3: Posterior Densities for Directions

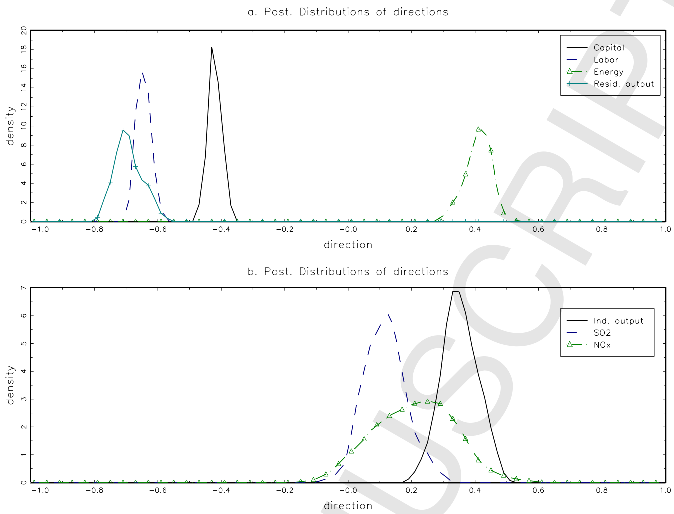

Fig. 4: Posterior Density for Technical Efficiency for Models I-III

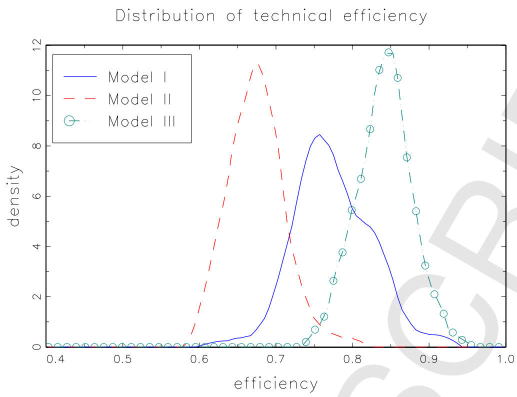

**Fig. 5: Posterior Densities for Price Distortions–
Good Inputs and Good Outputs**

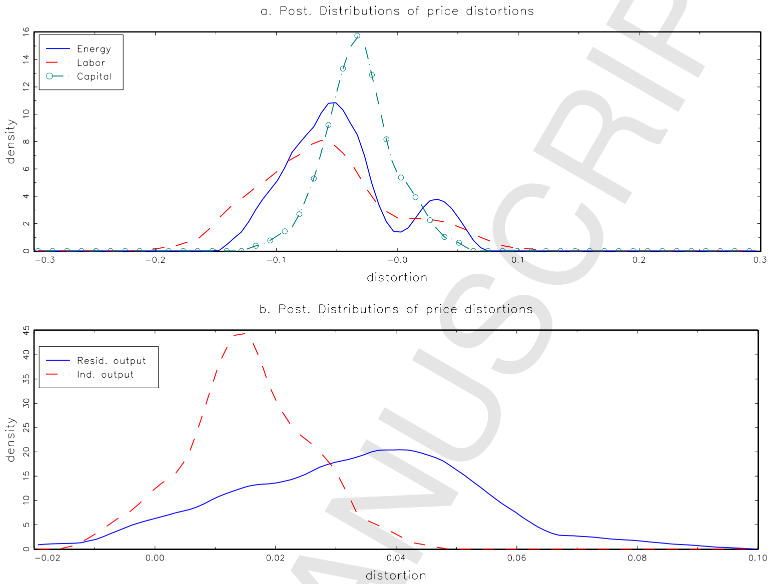

Fig. 6: Posterior Densities for Price Distortions–Bad Outputs

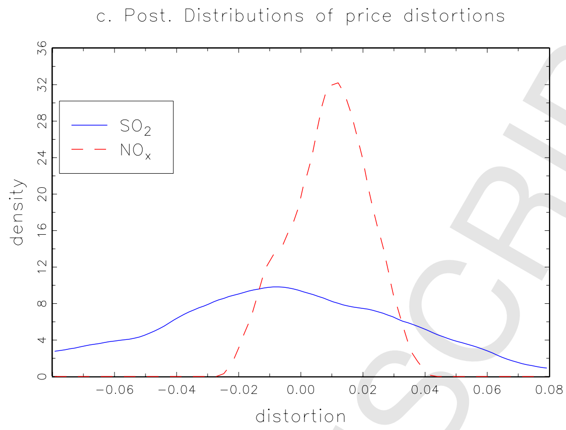

Fig. 7: Posterior Density for Allocative Efficiency for Models I-III

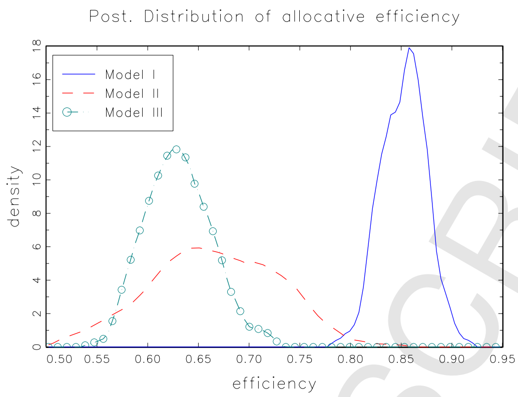

Fig. 8: Posterior Densities of Change in Inputs and Outputs

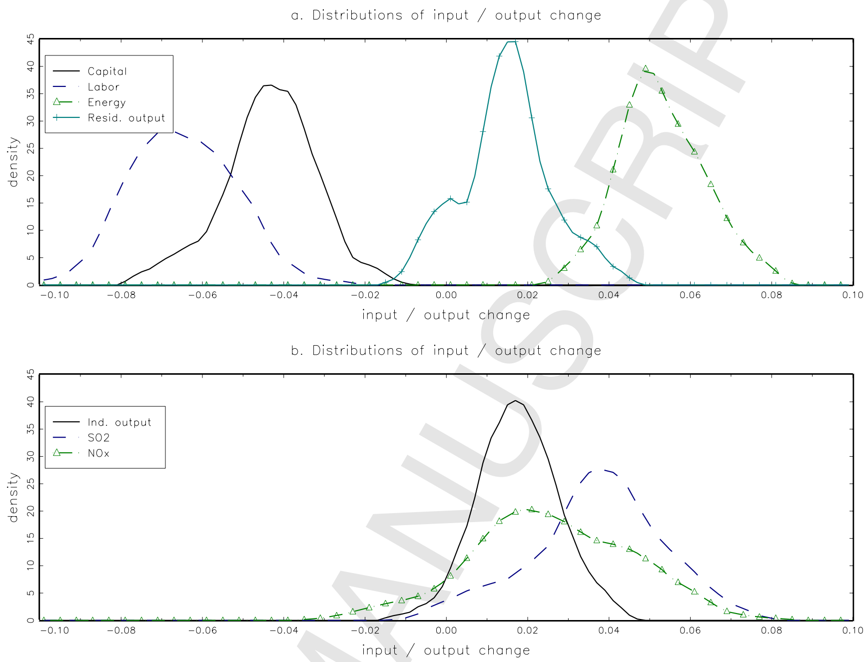

Fig. 9: Posterior Densities for RTS, PC, TC, and EC

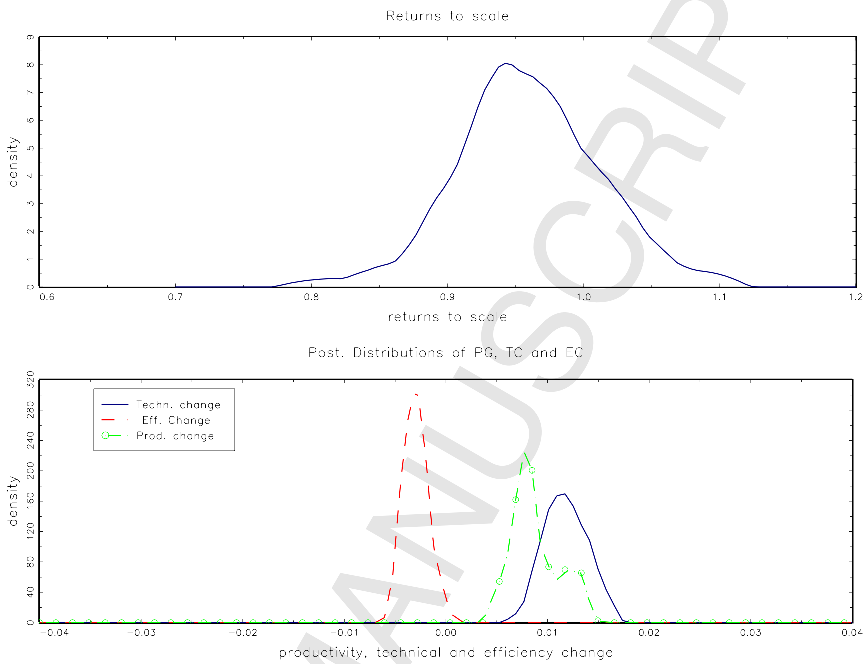

Fig. 10: Posterior Densities for Sensitivity of Changes in Inputs and Outputs to 10,000 Priors

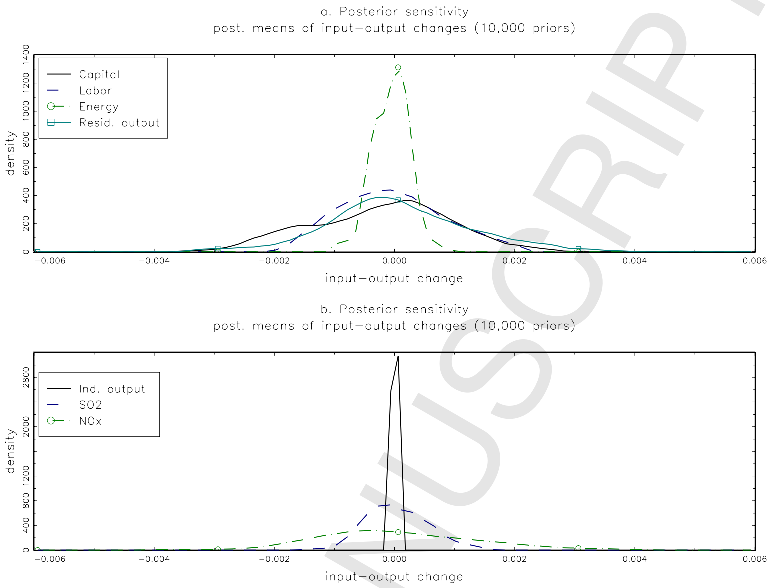

Fig. 11: Posterior Densities for Sensitivity of Changes in Structural Parameters to 10,000 Priors

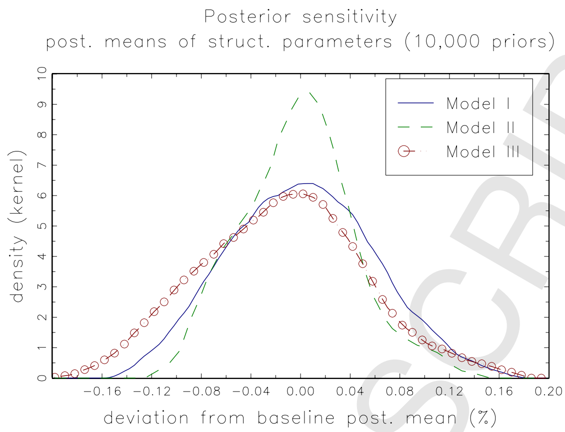

Fig. 12: Posterior Means of Relative Prices of Bad Outputs over Time from PM Model

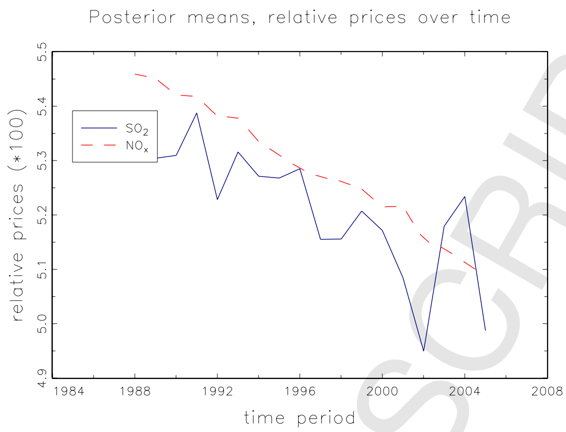

Fig. 13: Posterior Density of Predictive Measure, R, for PM Model

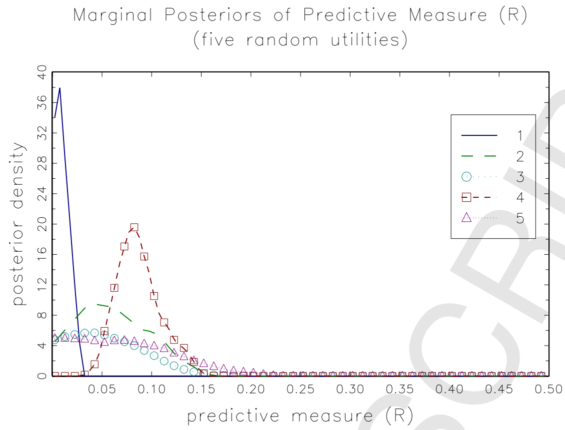

Fig. 14: Partial Elasticities for ω_{it} for the PM Model

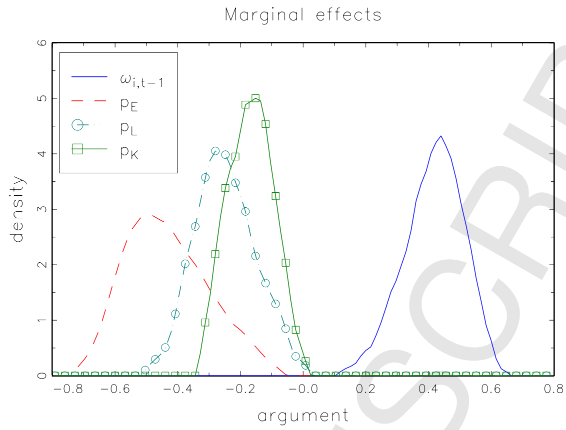

Fig. 15: Partial Elasticities for ω_{it} for the PM Model