

The Good, The Bad and The Technology: Endogeneity in Environmental Production Models*

Subal C. Kumbhakar[†]
Department of Economics
Binghamton University
Binghamton, New York

Efthymios G. Tsionas[‡]
Lancaster University Management School
Bailrigg LA1 4YX, UK

February 13, 2015

Abstract

In this paper we consider an environmental production process in which firms intend to produce outputs (which we label as desired/good outputs) but the production process is such that it automatically produces some other unintentional but inevitable undesirable (bad) outputs as by-products (emission of pollutants). Like stochastic frontier, by-production technology specifies that there is a certain minimal amount of the by-product that is produced, given the quantities of certain inputs and desired outputs. The presence of (environmental) inefficiency in by-production could generate more than this minimal amount of the undesired output. Similarly, the presence of technical inefficiency may imply that, given inputs, less than the maximal possible amount of desired output is produced. Alternatively it may imply that more than the minimal amount of inputs are used to produce a given level of desired output. We use the “by-production technology” approach which is a composition of production technology of desired outputs and the technology, and estimate both technical and environmental efficiencies. Given that electricity, the good output in our application, is demand determined, we treat it as exogenous and address the endogeneity of inputs by using the first-order conditions of cost minimization. Since we specify a separate technology for each bad output, their endogeneity is automatically taken into account. We use an efficient Bayesian MCMC technique to estimate the technologies and both types of inefficiencies. We also compare results with some alternative models with and without endogeneity corrections.

JEL Classification No.: C11, D21.

Keywords: By-production, Endogeneity, MCMC, Technical and Environmental Efficiency, Shadow Price.

*We thank Peter Schmidt and two anonymous referees for their constructive comments on an earlier version of the paper. We also thank Subhash Ray, the participants of Workshop on productivity analysis held at the University of Miami (Oct 25, 2013), and the seminar participants at the Department of Agricultural and Resource Economics, University of Connecticut, Storrs (Nov 1, 2013). Editorial help from Daniel Parisian and the Kumbhakar family members is highly appreciated.

[†]Corresponding author: Subal C. Kumbhakar, Department of Economics, State University of New York at Binghamton, PO Box 6000, Binghamton, NY 13902-6000. Phone: 607-777-4762, Fax: 607-777-2681. Email: kkar@binghamton.edu

[‡]Efthymios G. Tsionas, Lancaster University Management School, Bailrigg LA1 4YX UK. E-mail: tsionas@aueb.gr

1 Introduction

In this paper we consider an environmental production process in which firms intend to produce outputs (which we label as desired/good outputs) but the production process is such that it automatically produces some other undesirable (bad) outputs as by-products (emission of pollutant gases). Electric power generation is an example of a production process in which the production of electricity (desirable output) is accompanied by the emission of pollutants (undesirable outputs). This unintentional but inevitable outcome is labeled as “by-production” (BP). Like stochastic frontier models, BP technologies specify that there is a certain minimal amount of the undesired output that is produced, given the quantities of certain inputs and desired outputs. The presence of inefficiency in BP could generate more than this minimal amount of the undesired output. We call this environmental inefficiency. Similarly, the presence of technical inefficiency in the production of good output may imply that, given inputs, less than the maximal possible amount of desired output is produced. Alternatively it may imply that more than the minimal amount of inputs are used to produce a given level of desired output. We call this technical inefficiency.

The modeling of environmental production technology follows two distinct routes. In the directional distance function and transformation function approaches (see Färe et al. (2005), Atkinson and Dorfman (2005), Fernández, Koop and Steel (2005), Agee et al. (2014), among others), the technology is specified by a single equation in which good and bad outputs as well as good and bad (pollution generating) inputs enter as arguments. Inefficiency in the directional distance function approach is defined over the entire vector of outputs, both desirable and undesirable, using an *a priori* specified directional vector. Thus it confounds the technical inefficiency oriented along desirable outputs from the environmental inefficiency oriented along undesirable outputs. Similarly, inefficiency in the transformation function approach is either output or input oriented and it cannot be separated from environmental inefficiency. On the contrary, the BP approach is a composition of a production technology for good outputs and a residual-generation technology (production of bad outputs). Since the BP approach separates technology of production of good outputs from those of bad outputs, it can identify and estimate both technical and environmental inefficiency.

In this paper we use a series of models from both approaches and examine their suitability in capturing the features of environmental production technology. First, we consider a modeling approach that relies on the BP concept introduced in Murty, Russell and Levkoff (2012, MRL), Førsund (2009) and Fernández, Koop and Steel (2002, FKS). The technology for the production of good outputs is specified in terms of a standard translog transformation function (TF) with input-oriented technical inefficiency. In electricity generation, the arguments of the TF are good inputs (labor and capital), bad inputs (high sulfur coal) and good output (electricity). The technologies for the production of bad outputs (SO_2 and NO_x) are also specified as translog functions in which the arguments are good output and bad inputs. We consider two specifications. In (i) we use two separate technologies for two bad outputs, while in (ii) a single technology is assumed to produce both bad outputs. We estimate environmental efficiency for each bad output in (i) whereas in (ii) we estimate a single (radial) environmental efficiency for all bad outputs. In both specifications we correct for the endogeneity of inputs by explicitly using the first-order conditions of cost minimization (which is standard in the electricity generation literature going back to Nerlove (1965) in estimating the single as well as the multiple equation representation of the technology.

Second, since the BP approach is new and has not yet been econometrically estimated using a flexible functional form with monotonicity constraints, for comparison purpose we also consider standard single equation technology models similar to FKS (2005), Hailu and Veeman (2001), Agree (2014), among others. A translog TF, in which the arguments are both bad and good inputs as well as bad and good outputs, is used to represent it. Since the technology is specified by a single equation, only one type of inefficiency (which is a mixture of technical and environmental inefficiency) can be estimated. Further, there is no consensus on which variables

(good outputs, bad outputs, good inputs, bad inputs) are endogenous in this framework.

Third, we discuss pros and cons of each model and consider alternative versions of the single and system approaches. Finally, we compare results across different models, some of which correct for endogeneity while others do not.

The rest of the paper is organized as follows. Section 2 discusses various modeling approaches. Models with endogenous input in a cost minimizing framework is introduced in Section 3. Section 4 presents the data and Section 5 presents the empirical results. Section 6 concludes the paper. Details on the implementation of the MCMC technique can be found in the online appendix.

2 Modeling Technology without Endogeneity

In this section we consider models which do not address endogeneity issue at all. First, we consider the models in which the technology is specified by a single equation followed by the by-production approach.

2.1 Single Equation Production Technology: A Critique

In the single equation strand of the literature either a directional distance function (DDF) or a transformation function (TF) is used to represent the technology producing both good and bad outputs using both good and bad inputs. Here we consider the TF approach only and examine its merit from both theoretical and empirical points of view.¹

2.1.1 The Single Equation TF Formulation

The single equation model we consider here uses the TF (FKS, 2005; Atkinson and Dorfman, 2005) which can be represented by $F(Y, X^g, X^b, Z) = 1$, where Y, Z, X^g, X^b are vectors of M good outputs, Q bad outputs, J good inputs and K pollution generating (bad) inputs. The monotonicity assumptions on these variables are:² $F_Y \geq 0$, $F_{X^g} \leq 0$, $F_{X^b} \leq 0$ and $F_Z \leq 0$, where F_Y, F_{X^g}, F_{X^b} and F_Z are partial derivatives of $F(\cdot)$. Since $F_{X^g} \leq 0$, $F_{X^b} \leq 0$ and $F_Z \leq 0$, from a pure mathematical point of view, there is no difference between Z, X^g and X^b in $F(Y, X^g, X^b, Z) = 1$. That is, bad outputs can be treated as inputs³ (both X^g and X^b), and since inputs are assumed to be freely disposable so are bad outputs. This violates axioms of production theory and is criticized in the environmental production literature (for example, see Färe and Grosskopf (2003) and Färe et al. (2005)). Further, if in estimating the input distance function all bad outputs are treated as inputs, both bad outputs and inputs will be scaled back by the same proportion, holding good outputs constant. However, this approach violates a basic engineering requirement that a reduction in bad outputs requires the usage of more good inputs, holding good outputs and inefficiency constant.

There are some other problems associated with this model. For example, with 2 good outputs $\partial \ln Y_1 / \partial \ln Y_2 \leq 0$, ceteris paribus, since $F_{Y_1} \geq 0$ and $F_{Y_2} \geq 0$. This might be intuitive because when less of Y_1 is produced some resources will be released which can be used to increase the production of Y_2 . Similarly, for two bad outputs $\partial \ln Z_1 / \partial \ln Z_2 \leq 0$ if the monotonicity restrictions are imposed. However, from an engineering production point of view Z_1 and Z_2 might not be complementary. That is, more SO_2 emissions do not necessarily mean less of NO_x . The concave relationship between good and bad outputs means that when less of Z_1 is produced, less of Y is also produced which in turn means less of Z_2 , given the inputs and inefficiency. Thus $\partial \ln Z_1 / \partial \ln Z_2$ is expected to be positive (for an efficient unit at least), whereas it will be negative if the monotonicity restrictions are imposed.

¹For the DDF formulation the readers are advised to see Tsionas et al. (2014).

²For simplicity in exposition we treat Y, X^g, X^b, Z as scalars here and in the next three paragraphs.

³See Reinhard et al. (1999), Reinhard and Thijssen (2000), Lee et al. (2002), Hailu and Veeman (2001), among others.

Second, the concave relationship between Y and Z (Färe, 2005) is defined holding inputs constant. One might allow inputs to change and argue that a firm can reduce, say, Z (SO_2) by reducing X^b (for example, using high heat-content (or low-sulfur) coal in place of low heat-content (or high-sulfur) coal), holding Y and X^g unchanged.⁴ Taking the derivative of $F(Y, X^g, X^b, Z)$ holding Y and X^g constant gives $F_X^b dX^b + F_Z dZ = 0 \Rightarrow dZ/dX^b = -F_X^b/F_Z \leq 0$. This should be positive, i.e., SO_2 should be lower if less high-sulfur coal (X^b) is used. Thus we have a counter-intuitive result. This is also noted in MRL (2012, page 120, see their equation (2.8) and comment in item (b)).

Third, quite often we are interested in calculating the responsiveness in Z due to changes in X^b , X^g and Y (for example, Agee et al. (2014)). Note that unless there is inefficiency, BP implies that Z changes when X^b , X^g and Y change either individually or jointly. Thus, in calculating the responsiveness in Z due to changes in X^b , X^g or Y , the Z variable has to be related to either X^b , X^g , Y or all of them, and this relationship has to be used somehow in the model. For example, if $F(Y, X^g, X^b, Z) = 1$ is estimated in the form $X^g = f(Z, Y, X^b)$ which assumes X^g to be endogenous (and others predetermined/exogenous), one cannot compute $\partial Z/\partial X^b$, $\partial Z/\partial Y$ or even $\partial Z/\partial X$ from the estimated model.⁵ To compute $\partial Z/\partial X^b$, $\partial Z/\partial Y$, etc., one has to explicitly consider $Z = Z(X^b)$, $Z = Z(X^b, X^g)$ or $Z = Z(X^b, X^g, Y)$ and use them either as separate equations or embed them in $F(Y, X^g, X^b, Z)$ in some form. None of these are done in the single equation TF representation.

Since we use the single equation TF representation of the technology as one of our competing empirical models, we continue our discussion of the single equation TF model where Y, X^g, X^b and Z are vectors. We follow the TF formulation used in Caves et al. (1981) but extend it to accommodate input-oriented (IO) inefficiency (in good inputs) together with bad inputs and bad outputs. Thus our TF is

$$F(Y, \theta X^g, X^b, Z, t) = 1 \tag{1}$$

where as before $X^g \in \mathfrak{R}^J$ and $X^b \in \mathfrak{R}^K$ are the vectors of good and bad inputs, $Y \in \mathfrak{R}^M$ and $Z \in \mathfrak{R}^Q$ are vectors of good and bad outputs, t is a time trend (to capture technical change in a panel data setting) and $\theta \leq 1$ is input-oriented technical inefficiency. The TF $F(\cdot)$ is assumed to satisfy all the standard monotonicity properties (discussed later). We assume that good outputs Y are exogenously given (in our application the good output is electricity which is demand determined and therefore exogenous to the firm). This justifies an input distance function (IDF) representation of (1), which after imposing the linear homogeneity restrictions (in θX^g) can be written as $(\theta X_1^g)^{-1} = \psi(Y, \tilde{X}^g, X^b, Z, t)$ when $\tilde{X}^g = (X_2^g/X_1^g, \dots, X_J^g/X_1^g)$.⁶

Using $i = 1, \dots, n$ and $t = 1, \dots, T$ subscripts for firm and time in panel data setting, we write $\hat{X}_{jit}^g = \theta X_{jit}^g \Rightarrow \hat{x}_{jit}^g = x_{jit}^g + \ln \theta_{it}$ where the lowercase variables are in logarithmic form and assume a translog form of the transformation function, i.e.,

⁴For that to happen in the model $F(Y, X^g, X^b, Z) = 1$, one has to explicitly consider $Z = Z(X^b)$ either as a separate equation or embed it in $F(Y, X^g, X^b, Z)$ in some form.

⁵For example, if the model $Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + u$ is estimated in a way that takes endogeneity of, for example, X_1 into account, can one interpret $\partial X_1/\partial X_2$ in a meaningful way?

⁶Note that although linear homogeneity helps one to write the TF function in the form of an IDF $(\theta X_1^g)^{-1} = \psi(Y, \tilde{X}^g, X^b, Z, t)$, it does not imply that the covariates (Y, \tilde{X}^g, X^b, Z) are exogenous/predetermined. However, this is routinely assumed in estimating the IDF.

$$\begin{aligned}
\ln F(Y_{it}, \theta_{it} X_{it}^g, X_{it}^b, Z_{it}, t) = & \alpha_o + \sum_{j=2}^J \alpha_j \hat{x}_{j,it}^g + \sum_{m=1}^M \beta_m y_{m,it} + \frac{1}{2} \sum_{j=1}^J \sum_{j'=1}^J \alpha_{jj'} \hat{x}_{j,it}^g \hat{x}_{j',it}^g + \alpha_t t + \frac{1}{2} \alpha_{tt} t^2 + \sum_m \beta_{mt} y_{m,it} t \\
& + \frac{1}{2} \sum_{m=1}^M \sum_{m'=1}^M \beta_{mm'} y_{m,it} y_{m',it} + \sum_{j=1}^J \sum_{m=1}^M \delta_{jm} \hat{x}_{j,it}^g y_{m,it} + \sum_j \alpha_{jt} \hat{x}_{j,it}^g t + \sum_{k=1}^K \tilde{\alpha}_k x_{k,it}^b \\
& + \frac{1}{2} \sum_{q=1}^Q \sum_{q'=1}^Q \gamma_{qq'} z_{q,it} z_{q',it} + \sum_{q=1}^Q \sum_{m=1}^M \bar{\eta}_{jm} \hat{x}_{j,it}^g z_{q,it} + \sum_{q=1}^Q \sum_{m=1}^M \chi_{jm} z_{j,it}^q y_{m,it} + \sum_{q=1}^Q \bar{\chi}_{qt} z_{q,it} t \\
& + \frac{1}{2} \sum_{k=1}^K \sum_{k'=1}^K \zeta_{kk'} x_{k,it}^b x_{k',it}^b + \sum_{j=1}^J \sum_{k=1}^K \varphi_{jm} \hat{x}_{j,it}^g x_{k,it}^b + \sum_{k=1}^K \sum_{m=1}^M \psi_{jm} x_{k,it}^b y_{m,it} \\
& + \sum_{k=1}^K \bar{\psi}_{jt} x_{k,it}^b t. \tag{2}
\end{aligned}$$

Upon imposing linear homogeneity restrictions, the above translog function can be written as⁷

$$\begin{aligned}
x_{1,it}^g = & \alpha_o + \sum_{j=2}^J \alpha_j \tilde{x}_{j,it}^g + \sum_{m=1}^M \beta_m y_{m,it} + \frac{1}{2} \sum_{j=2}^J \sum_{j'=2}^J \alpha_{jj'} \tilde{x}_{j,it}^g \tilde{x}_{j',it}^g + \alpha_t t + \frac{1}{2} \alpha_{tt} t^2 + \sum_m \beta_{mt} y_{m,it} t \\
& + \frac{1}{2} \sum_{m=1}^M \sum_{m'=1}^M \beta_{mm'} y_{m,it} y_{m',it} + \sum_{j=2}^J \sum_{m=1}^M \delta_{jm} \tilde{x}_{j,it}^g y_{m,it} + \sum_j \alpha_{jt} \tilde{x}_{j,it}^g t \\
& + \frac{1}{2} \sum_{q=1}^Q \sum_{q'=1}^Q \gamma_{qq'} z_{q,it} z_{q',it} + \sum_{q=1}^Q \sum_{m=1}^M \bar{\eta}_{jm} \tilde{x}_{j,it}^g z_{q,it} + \sum_{q=1}^Q \sum_{m=1}^M \chi_{jm} z_{j,it}^q y_{m,it} + \sum_{q=1}^Q \bar{\chi}_{qt} z_{q,it} t \\
& + \frac{1}{2} \sum_{k=1}^K \sum_{k'=1}^K \zeta_{kk'} x_{k,it}^b x_{k',it}^b + \sum_{j=2}^J \sum_{k=1}^K \varphi_{jm} \tilde{x}_{j,it}^g x_{k,it}^b + \sum_{k=1}^K \sum_{m=1}^M \psi_{jm} x_{k,it}^b y_{m,it} + \sum_j \alpha_{jt} \tilde{x}_{j,it}^g t + \sum_{k=1}^K \tilde{\alpha}_k x_{k,it}^b \\
& + \sum_{k=1}^K \bar{\psi}_{jt} x_{k,it}^b t + v_{1,it} + u_{1,it} \equiv TL(\tilde{x}_{it}^g, x_{it}^b, y_{it}, z_{q,it}, t) + v_{1,it} + u_{1,it}, \tag{3}
\end{aligned}$$

where, $TL(\cdot)$ represents a translog function with arguments inside the parentheses. As before, $\tilde{x}_{j,it}^g = x_{j,it}^g - x_{1,it}^g, j = 2, \dots, J$ and $u_{1,it} = -\ln \theta_{it} \geq 0$. We also add a stochastic noise term $v_{1,it} \geq 0$ in (3). One can estimate (3) using a standard stochastic cost frontier approach, assuming that $\tilde{x}_{j,it}^g, y_{m,it}, z_{q,it}$ and $x_{k,it}^b$ are exogenous/predetermined and uncorrelated with $v_{1,it}$ and $u_{1,it}$. This is a bold assumption and it is unlikely to be true. However, assuming that this assumption holds, one can estimate technical inefficiency by estimating (3). Note that there is no environmental inefficiency in this model.

2.2 A By-production Technology when Bad Outputs are Separately Produced

Given the problems in modeling bad outputs in terms of a single equation distance/TF, it is clear that we need separate tools to model production processes involving bad outputs. The salient feature of modeling environmental (pollution-generating) production technology is to incorporate the positive correlation between bad and good outputs. In a standard single equation representation of the technology (in terms of a distance

⁷Strictly speaking, there should be a minus sign in front of $x_{1,it}^g$. To remove the negative sign we multiply both sides of (7) by -1 which changes the sign of all the coefficients and the inefficiency term $u_{1,it}$ in (7). Note that we have not changed the signs on the parameters, these changes will be automatically absorbed by the estimated parameters.

or transformation function) this positive correlation is expected to be satisfied via the monotonicity restrictions. However, as noted above, imposition of the monotonicity conditions do not automatically satisfy all the standard axioms of production theory. Next we examine whether the BP approach can solve the problems discussed above.

Following FKS (2002), Førsund (2009), and MRL (2012), we consider the BP approach which uses two separate technologies to model good and bad outputs. The former describes the textbook type production process, i.e., inputs (good and bad) are transformed into desired outputs, and it does not depend on bad outputs. Further, it satisfies all the standard properties, most importantly the free-disposability property. The latter can be viewed as a residual generation technology which models the production of bad outputs as a function of either good outputs (FKS, 2002), bad inputs (MRL, 2012) or both bad and good inputs (Førsund, 2009). The positive relationship between bad and good outputs in FKS (2002) follow from this residual technology which embeds the relationship explicitly. Further, inefficiency is allowed in each technology, thereby distinguishing technical inefficiency from environmental inefficiency.

As mentioned in the introduction, we consider two separate BP technology models. First, we assume that the technology for production of each bad output is different. This approach can be justified provided that, on engineering grounds, there is no substitutability/complementarity between bad outputs although the same polluting input (high sulfur, low heat-content coal) can produce more of the bad outputs (SO_2 and NO_x). We also consider a model in which a single technology produces all bad outputs as in FKS (2002). This model will allow substitutability/complementarity between bad outputs. In both cases the technology of good outputs is the same. We use Bayes factor to test which model performs best in predictive terms.

As before, we start with the TF representation of the underlying technology with IO inefficiency for the production of good outputs, viz.,

$$F(Y, \theta X^g, X^b, t) = 1, \quad (4)$$

Note that bad output vector Z does not appear in (4). The transformation function $F(\cdot)$ is assumed to satisfy all the standard monotonicity properties (discussed later). Using the linear homogeneity restrictions (in θX^g) the transformation function in (4) can be expressed as $(\theta X_1^g)^{-1} = \psi^g(Y, \tilde{X}^g, X^b, t)$.

In specifying the technology for the production of bad outputs, we consider two models. In Model 1 we allow separate technology for the production of each bad output, i.e., $H_q(Y, \lambda_q Z_q, X^b, t) = 1$ where $\lambda_q \leq 1$ is environmental inefficiency in the production of Z_q . More specifically, $(1 - \lambda_q)$ 100% is the rate at which the production of bad output Z_q can be reduced without reducing good outputs, and bad inputs. Thus in Model 1 we do not allow substitutability among bad outputs.

The technologies for the production of bad outputs $H_q(Y, \lambda_q Z_q, X^b, t) = 1$ are rewritten as:

$$z_{q,it} = g_q(y_{it}, x_{it}^b, t) + \xi_{q,it} + \eta_{q,it}, \quad q = 1, \dots, Q \quad (5)$$

where $z_{q,it}$ represents the (log of) bad output Z_q ($q = 1, \dots, Q$). Furthermore, $\eta_{q,it} = -\ln \lambda_q \geq 0$ represents environmental inefficiency in the sense that it gives the percentage over production of Z_q , *ceteris paribus*. Finally, we added a stochastic error, $\xi_{q,it} \lesseqgtr 0$ (similar to $v_{j,it}$), for each q .

As before we write $\hat{X}_{jit}^g = \theta_{it} X_{jit} \Rightarrow \hat{x}_{jit}^g = x_{jit}^g + \ln \theta_{it}$ and assume a translog form of the TF in (4), i.e.,

$$\begin{aligned}
\ln F(Y_{it}, \theta_{it} X_{it}^g, X_{it}^b, t) &= \alpha_o + \sum_{j=1}^J \alpha_j \hat{x}_{j,it} + \sum_{m=1}^M \beta_m y_{m,it} + \frac{1}{2} \sum_{j=1}^J \sum_{j'=1}^J \alpha_{jj'} \hat{x}_{j,it} \hat{x}_{j',it} + \alpha_t t \\
&+ \frac{1}{2} \alpha_{tt} t^2 + \sum_m \beta_{mt} y_{m,it} t + \frac{1}{2} \sum_{m=1}^M \sum_{m'=1}^M \beta_{mm'} y_{m,it} y_{m',it} + \sum_j \alpha_{jt} \hat{x}_{j,it} t + \sum_{k=1}^K \tilde{\alpha}_k x_{k,it}^b \\
&+ \sum_{j=1}^J \sum_{m=1}^M \delta_{jm} \hat{x}_{j,it} y_{m,it} + \sum_j \alpha_{jt} \hat{x}_{j,it} t + \frac{1}{2} \sum_{k=1}^K \sum_{k'=1}^K \zeta_{kk'} x_{k,it}^b x_{k',it}^b \\
&+ \sum_{j=1}^J \sum_{k=1}^K \varphi_{jm} \hat{x}_{j,it} x_{k,it}^b + \sum_{k=1}^K \sum_{m=1}^M \psi_{jm} x_{k,it}^b y_{m,it} + \sum_{k=1}^K \bar{\psi}_{jt} x_{k,it}^b t. \tag{6}
\end{aligned}$$

We impose linear homogeneity with respect to θX^g to the above transformation function, and rewrite it as

$$\begin{aligned}
x_{1,it}^g &= \alpha_o + \sum_{j=2}^J \alpha_j \tilde{x}_{j,it}^g + \sum_{m=1}^M \beta_m y_{m,it} + \frac{1}{2} \sum_{j=2}^J \sum_{j'=2}^J \alpha_{jj'} \tilde{x}_{j,it}^g \tilde{x}_{j',it}^g + \alpha_t t + \frac{1}{2} \alpha_{tt} t^2 + \sum_m \beta_{mt} y_{m,it} t \\
&+ \frac{1}{2} \sum_{m=1}^M \sum_{m'=1}^M \beta_{mm'} y_{m,it} y_{m',it} + \sum_{j=2}^J \sum_{m=1}^M \delta_{jm} \tilde{x}_{j,it}^g y_{m,it} + \sum_j \alpha_{jt} \tilde{x}_{j,it}^g t + \frac{1}{2} \sum_{k=1}^K \sum_{k'=1}^K \zeta_{kk'} x_{k,it}^b x_{k',it}^b \\
&+ \sum_{k=1}^K \tilde{\alpha}_k x_{k,it}^b + \sum_{j=1}^J \sum_{k=1}^K \varphi_{jm} \tilde{x}_{j,it}^g x_{k,it}^b + \sum_{k=1}^K \sum_{m=1}^M \psi_{jm} x_{k,it}^b y_{m,it} + \sum_{k=1}^K \bar{\psi}_{jt} x_{k,it}^b t + v_{1,it} + u_{1,it} \\
&\equiv TL(\tilde{x}_{j,it}, y_{m,it}, t) + v_{1,it} + u_{1,it} = \alpha_o + \mathbf{S}'_{it} \mathbf{a} + \frac{1}{2} \mathbf{S}'_{it} \mathbf{A} \mathbf{S}_{it} + v_{1,it} + u_{1,it}, \tag{7}
\end{aligned}$$

where $\tilde{x}_{j,it}^g = x_{j,it}^g - x_{1,it}^g, j = 2, \dots, J$, $\mathbf{S}'_{it} = (y'_{it}, x'_{it}, \tilde{x}'_{it}, t)$, $u_{1,it} = -\ln \theta_{it} \geq 0$, and \mathbf{a} , \mathbf{A} are vector and matrix of relevant parameters. We also add a stochastic noise term $v_{1,it} \geq 0$ in (7).

We assume translog functional forms on $g_q(\cdot)$ in (5), and write it more compactly as,

$$z_{q,it} = a_{q,0} + \mathbf{R}'_{it} \boldsymbol{\beta}_q + \frac{1}{2} \mathbf{R}'_{it} \boldsymbol{\Gamma}_q \mathbf{R}_{it} + \xi_{q,it} + \eta_{q,it}, \quad q = 1, \dots, Q, \tag{8}$$

where $\mathbf{R}'_{it} = (y'_{it}, x'_{it}, t)$, and $\boldsymbol{\beta}_q, \boldsymbol{\Gamma}_q$ are vectors and matrices of relevant parameters in each of the translog function representing production of bad output. Thus Model 1 consists of the technologies for the production of good and bad outputs in (7) and (8).

2.3 A By-production Model when Bad Outputs are Jointly Produced

Now we consider an alternative model (Model 2) in the spirit of FKS (2002), Førsund (2009), and MRL (2012). The technology for good output is specified in terms of the TF in (7) is exactly the same that is used in Model 1. However, in Model 2, there is only one technology for the production of bad outputs, i.e., $H(Y, \lambda Z, X^b, t) = 1$ where $\lambda \leq 1$ is environmental inefficiency in the production of all bad outputs. This radial measure shows that production of all bad outputs Z can be reduced by $(1 - \lambda)$ 100% without reducing good outputs and bad inputs. Note that in Model 2 we use radial measures for both technical and environmental inefficiency. In contrast, Model 1 uses radial measure of technical inefficiency but non-radial measure of environmental inefficiency.

We assume that $H(Y, \lambda Z, X^b, t) = 1$ is homogeneous of degree 1 in Z .⁸ Using the homogeneity assumption,

⁸FKS (2002) did not distinguish between good and bad inputs and did not use them in the BP function $H(\cdot)$. Furthermore, they used a more restrictive form of $H(\cdot)$, viz., $H(Y, \lambda Z) = h(Y) \cdot g(\lambda Z)$.

we can rewrite $H(Y, \lambda Z, X^b, t) = 1$ as $(\lambda Z_1)^{-1} = h(Y, \tilde{Z}, X^b, t)$ where $\tilde{Z} = (Z_2/Z_1, \dots, Z_Q/Z_1)$

Thus, in Model 2 the technology is represented by the following two equations, i.e., $(\theta X_1^g)^{-1} = \psi^g(Y, \tilde{X}^g, X^b, t)$ and $(\lambda Z_1)^{-1} = h(Y, \tilde{Z}, X^b, t)$. Assuming translog functional form on both, Model 2 is specified as

$$x_{1,it}^g = \alpha_0 + \mathbf{S}'_{it} \mathbf{a} + \frac{1}{2} \mathbf{S}'_{it} \mathbf{A} \mathbf{S}_{it} + v_{1,it} + u_{1,it}, \quad (9)$$

$$z_{1,it} = a_0 + \mathbf{P}'_{it} \boldsymbol{\delta} + \frac{1}{2} \mathbf{P}'_{it} \boldsymbol{\Delta} \mathbf{P}_{it} + \zeta_{it} + \tau_{it}, \quad (10)$$

where $\mathbf{P}'_{it} = (y'_{it}, \tilde{z}'_{it}, x_{it}^b, t)$, $\tilde{z}_{qit} = z_{qit} - z_{1it}$, $\tau_{it} = \ln \lambda_{it} \geq 0$ and ζ_{it} is an error term. Further, $\boldsymbol{\delta}$ and $\boldsymbol{\Delta}$ are vector and matrix of relevant parameters in the translog function representing production of bad output. The other parameter vector and matrices \mathbf{a} and \mathbf{A} are already defined. Note that (9) is the same as (7), written in a compact form.

Both the models should satisfy the monotonicity conditions⁹ (standard properties of an IDF). These are:

$$\frac{\partial x_{1,it}^g}{\partial \tilde{x}_{j,it}^g} \leq 0, j = 2, \dots, J \quad (11)$$

$$\frac{\partial x_{1,it}^g}{\partial y_{m,it}} \geq 0, m = 1, \dots, M \quad (12)$$

$$\frac{\partial x_{1,it}^g}{\partial x_{k,it}^b} \leq 0, k = 1, \dots, K \quad (13)$$

The exact form of the above derivatives can be obtained by differentiating the TF in (7) with respect to $\tilde{x}_{j,it}^g$, $y_{m,it}$ and $x_{k,it}^b$.

Similarly to guarantee a positive relationship between each pair of good and bad outputs, as well as each pair of bad output and bad inputs, we need to impose the following restrictions in Model 1:

$$\frac{\partial z_{q,it}}{\partial y_{m,it}} \geq 0, m = 1, \dots, M \quad (14)$$

$$\frac{\partial z_{q,it}}{\partial \tilde{x}_{k,it}^b} \geq 0, k = 1, \dots, K \quad (15)$$

for each $q = 1, \dots, Q$. The exact form of the above derivatives can be obtained by differentiating (8) with respect to $y_{m,it}$ and $\tilde{x}_{k,it}^b$.

For Model 2, these restrictions are

$$\frac{\partial z_{1,it}}{\partial y_{m,it}} \geq 0, m = 1, \dots, M \quad (16)$$

$$\frac{\partial z_{1,it}}{\partial \tilde{x}_{k,it}^b} \geq 0, k = 1, \dots, K \quad (17)$$

the exact expression of which can be obtained by differentiating the equation in (10).

The models we have considered so far are summarized in Table 1. Model 0 uses a single TL IDF to represent the technology for the production of good and bad outputs in which x_1^g is endogenous.¹⁰ Model 1 has one TL IDF that describes the production of good outputs, and one TL function for each z_q to represent the technology for the production of each bad output. Thus, the endogenous variables in Model 1 are x_1^g and $\mathbf{z} = \{z_q, q = 1, \dots, Q\}$.

⁹The signs are reversed because we have x_1 instead of $-x_1$ on the left-hand-side of the IDF

¹⁰Note that the lowercase variables are in logs.

Model 2 has one TL IDF that describes the production of good outputs and one TL function that represents the technology for the production of all bad outputs. The endogenous variables in this model are x_1^g and z_1 .

Table 1: Description of Models

Model	Description	Endogenous	Predetermined/ Exogenous
Model 0	$x_1^g = TL(\tilde{\mathbf{x}}^g, \mathbf{x}^b, \mathbf{y}, \mathbf{z}, t)$	x_1^g	$\mathbf{y}, \tilde{\mathbf{x}}^g, \mathbf{x}^b, \mathbf{z}, t$
Model 1	$x_1^g = TL(\tilde{\mathbf{x}}^g, \mathbf{x}^b, \mathbf{y}, t)$, $\mathbf{z} = \mathbf{g}(\mathbf{y}, \mathbf{x}^b, t)$ (one TL function for each z)	x_1^g, \mathbf{z}	$\mathbf{y}, \tilde{\mathbf{x}}^g, \mathbf{x}^b, t$
Model 2	$x_1^g = TL(\tilde{\mathbf{x}}^g, \mathbf{x}^b, \mathbf{y}, t)$, $z_1 = TL(\mathbf{y}, \mathbf{x}^b, \tilde{\mathbf{z}})$ (one TL function for z_1)	x_1^g, z_1	$\mathbf{y}, \tilde{\mathbf{x}}^g, \mathbf{x}^b, \tilde{\mathbf{z}}, t$

3 Modeling Technology when Good Inputs are Endogenous

We now consider the same models (Models 0-2) but use explicit cost minimization behavior to address the endogeneity of good inputs. That is, the first-order conditions of cost minimization are now used along with the technologies of producing good and bad outputs.

3.1 A Single Production Technology

We assume that producers choose X^g to minimize cost subject to the TF to produce a given level of Y , viz.,

$$\min_{X^g \in \mathfrak{R}^J} : W'X^g \text{ subject to } F(Y, \theta X^g, X^b, Z, t) = 1. \quad (18)$$

where W is the exogenously given vector of good input prices.

The first-order conditions (FOCs) of the above problem can be written as:

$$\frac{W_j}{W_1} = \frac{\partial F(Y, \theta X^g, X^b, Z, t) / \partial X_j^g}{\partial F(Y, \theta X^g, X^b, Z, t) / \partial X_1^g} \Rightarrow \frac{W_j X_j^g}{W_1 X_1^g} = \frac{\partial \ln F(Y, \theta X^g, X^b, Z, t) / \partial \ln X_j^g}{\partial \ln F(Y, \theta X^g, X^b, Z, t) / \partial \ln X_1^g}, \quad j = 2, \dots, J. \quad (19)$$

Define $S_{j,it} = \frac{\partial \ln F(Y, \theta X^g, X^b, Z, t)}{\partial x_{j,it}^g} = \frac{\partial x_{1,it}^g}{\partial \tilde{x}_{j,it}^g}$ and use (3) to get

$$S_{j,it} = \alpha_j + \sum_{j'=2}^J \alpha_{jj'} \tilde{x}_{j',it}^g + \sum_{m=1}^M \delta_{jm} y_{m,it} + \alpha_{jt} t + \sum_{q=1}^Q \bar{\eta}_{jm} z_{q,it} + \sum_{k=1}^K \varphi_{jm} x_{k,it}^b, \quad j = 2, \dots, J \quad (20)$$

and rewrite the FOCs in (19) as

$$\tilde{w}_{j,it} \equiv w_{j,it} - w_{1,it} = \ln [S_{j,it}/S_{1,it}] - x_{j,it}^g + x_{1,it}^g + v_{j,it}, \quad j = 2, \dots, J. \quad (21)$$

Since $\sum_{j=1}^J S_{j,it} = 1$, $S_{1,it} = 1 - \sum_{j=2}^J S_{j,it}$ where $S_{j,it}, j = 2, \dots, J$ comes from (20). The system consisting of (3) and (21) can be estimated assuming $(v_{1,it}, \dots, v_{J,it})'$ is distributed as multivariate normal and $u_{1,it}$ is half normal. Note that Y_m and Z_q are assumed to be exogenous.

3.2 A By-production Technology

The cost minimization problem for the BP technology model is:

$$\min_{X^g \in \mathfrak{R}^J} : W'X^g \text{ subject to } F(Y, \theta X^g, X^b, t) = 1. \quad (22)$$

The FOCs of the above problem are:

$$\frac{W_j}{W_1} = \frac{\partial F(Y, \theta X^g, X^b, t) / \partial X_j^g}{\partial F(Y, \theta X^g, X^b, t) / \partial X_1^g} \Rightarrow \frac{W_j X_j^g}{W_1 X_1^g} = \frac{\partial \ln F(Y, \theta X^g, X^b, t) / \partial \ln X_j^g}{\partial \ln F(Y, \theta X^g, X^b, t) / \partial \ln X_1^g}, j = 2, \dots, J. \quad (23)$$

The above FOCs in (23), after taking logarithms, can be rewritten as

$$\tilde{w}_{j,it} \equiv w_{j,it} - w_{1,it} = \ln [S_{j,it}^* / S_{1,it}^*] - x_{j,it}^g + x_{1,it}^g + v_{j,it}, j = 2, \dots, J, \quad (24)$$

where $S_{j,it}^* = \frac{\partial \ln F(Y, \theta X^g, X^b, t)}{\partial x_{j,it}^g} = \frac{\partial \ln x_{1,it}^g}{\partial x_{j,it}^g}$ which can be obtained from (7), viz.,

$$S_{j,it}^* = \alpha_j + \sum_{j'=2}^J \alpha_{jj'} \tilde{x}_{j',it}^g + \sum_{m=1}^M \delta_{jm} y_{m,it} + \sum_{k=1}^K \varphi_{jm} x_{k,it}^b + \alpha_{jt} t, j = 2, \dots, J. \quad (25)$$

Note that the difference between $S_{j,it}$ and $S_{j,it}^*$ in (20) and (25) comes from the fact the single equation technology has Z as arguments while the BP technology has no Z in the production of good outputs. If the technology for each bad output is different (Model 1), then the system of equations to be estimated jointly is given by (7), (24), and (8). Note that the endogenous variables in the system are x_j^g and z_q .

On the other hand, if a single technology produces all bad outputs (Model 2), then the system of equations to be estimated jointly is given by (7), (24), and (10). Note that the technology for the production of bad outputs is assumed to be homogeneous of degree 1 in Z . The endogenous variables in the system are $X_j^g, j = 2, \dots, J$ and Z_1 . Similar to the standard IDF (in which input ratios are argued to be predetermined), it is also assumed the ratios of bad outputs are predetermined/exogenous and are uncorrelated with the error components. This assumption is not required if the technology for each bad output is different (as in (8)).

The models discussed in the preceding section are summarized in Table 2. Model 0a uses a system consisting of the TL IDF that represent the technology for the production of good and bad outputs, and the FOCs with respect to X^g . Thus, the endogenous variables in Model 0a are $X_j, j = 1, \dots, J$. Similarly, the system in Model 1a consists of the IDF, the FOCs with respect to X_j^g and one TL function for each Z_q . Thus the endogenous variables are X_j^g and $Z_q, q = 1, \dots, Q$. Finally, the system in Model 2a consists of the IDF, the FOCs with respect to X_j^g and one TL function for Z_1 . Thus the endogenous variables are $X_j^g, j = 1, \dots, J$ and Z_1 .¹¹

4 Data

The data we use come from Pasurka (2006) and Murty et al. (2012). A balanced panel consisting of 92 coal-fired electric power generating plants operating in the U.S. over the period from 1985 to 1995 is used in this study. This gave us a total of 1,012 observations. We focus on coal-fired plants (with at least 95% of total fuel consumption (measured in Btu) from coal). The specification of outputs and inputs is as follows. The good output is the net electric power generation Y , measured in kWh. The two bad outputs are (i) the SO_2 (sulfur dioxide) gas emissions (Z_1) and (ii) the NO_x (nitrogen oxides) gas emissions (Z_2), both measured in short-tons.

¹¹Table 2 lists the variables in lowercase because all the variables in Models 0a, 1a, and 2a appear in log.

Table 2: Description of Models when Inputs are Endogenous

Model	Description	Endogenous	Predetermined/ Exogenous
Model 0a	$x_1^g = TL(\tilde{\mathbf{x}}^g, \mathbf{x}^b, \mathbf{y}, \mathbf{z}, t)$, FOCs in X_j^g	\mathbf{x}^g	$\mathbf{y}, \tilde{\mathbf{x}}^g, \mathbf{x}^b, \mathbf{z}, \tilde{\mathbf{w}}, t$
Model 1a	$x_1^g = TL(\tilde{\mathbf{x}}^g, \mathbf{x}^b, \mathbf{y}, t)$, FOCs in X_j^g , $z_q = g_q(\mathbf{y}, \mathbf{x}^b, t)$ (one TL function for each z_q)	\mathbf{x}^g, \mathbf{z}	$\mathbf{y}, \mathbf{x}^b, \tilde{\mathbf{w}}, t$
Model 2a	$x_1^g = TL(\tilde{\mathbf{x}}^g, \mathbf{x}^b, \mathbf{y}, t)$, FOCs in X_j^g , $z_1 = TL(\mathbf{y}, \mathbf{x}^b, \tilde{\mathbf{z}}, t)$ (only one TL function for z_1)	\mathbf{x}^g, z_1	$\mathbf{y}, \mathbf{x}^b, \tilde{\mathbf{z}}, \tilde{\mathbf{w}}, t$

The good inputs to the production are: the real stock of physical capital (X_1^g), constructed from historical cost of plant data and deflated to constant dollars using the Handy-Whitman Index; and labor (X_2^g), measured in the number of employees. The bad input is the heat content of coal (X^b), measured in Btu per ton of coal. This is proxy for quality of coal. If the quality of coal is low, the heat content will be low. The data on the cost of plants and equipment (used in the construction of the capital stock) and the number of employees come from the U.S. Federal Energy Regulatory Commission Form 1 survey. The data on fuel consumption, net power generation and pollutant gas emissions come from the U.S. Department of Energy Form EIA-767 survey. 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, obtained from FERC Form 1. The price of capital is the product of the yield of the firm's latest issue of long-term debt, adjusted for appreciation and depreciation of capital using the Christensen-Jorgenson (1970) cost of capital formula, and the Handy-Whitman index for electric utility construction costs.

5 Estimation

The models we discussed so far are broadly classified into two groups, a single technology and a BP technology. Models in each group are estimated with and without addressing endogeneity. Although X^g is viewed as endogenous in both Models 0 and 0a, there is a difference. In Model 0 inputs X_j^g are endogenous but $\tilde{X}_j^g = X_j^g/X_1^g, j = 2, \dots, J$ are treated as exogenous. Thus in fact, only X_1^g is treated as endogenous in Model 0 in which only the IDF is estimated. On the other hand, in Model 0a all the good inputs X_j^g are treated as endogenous because in addition to the IDF we also add $(J - 1)$ FOCs so that we have a system of J equations for J endogenous variables (X_j^g). In both Models 0 and 0a bad outputs are treated as exogenous. Similar differences are also in Models 1 and 1a and Models 2 and 2a. In Models 2 and 2a are only Z_1 is endogeneous and $\tilde{Z}_q = Z_q/Z_1, q = 2, \dots, Q$ are treated as exogenous. This is, however, not the case in Models 1 and 1a in which $X_j^g, j = 1, \dots, J$ and $Z_q, q = 1, \dots, Q$ are endogenous.

In estimation, Model 0 uses only the IDF in (3). Assuming that none of the covariates (good and bad outputs, bad inputs and the ratios of good inputs) are correlated with either the inefficiency or the noise term, estimation can be done using the standard stochastic cost frontier approach. If some of the covariates are endogenous, the estimates are likely to be inconsistent. Model 0a consists of the IDF in (3) and FOCs in (21). By taking the FOCs into account Model 0 allows X_j^g to be endogenous, not just X_1^g .

Models 1 and 2 use the BP approach. The only difference between these two models is that in Model 2 we assume a single production process for bad outputs, whereas in Model 1 we allow the technologies for the production of bad outputs to be different. If the monotonicity constraints are ignored then these models can be

estimated equation by equation using the ML method, routinely used in stochastic frontier cost models.

To summarize: Model 1 consists of (7) and (8), whereas Model 2 consists of (7) and (10). Models 1a consists of (7), (24) and (8). Model 2a consists of (7), (24) and (10). All these models are estimated with the appropriate monotonicity constraints imposed. Distributional assumptions are made on the noise and inefficiency components.

Consider the BP approach in Model 1a. We write the error terms of this system in vector form as $\mathbf{v}_{it} = [v_{1,it}, \dots, v_{J,it}, \xi_{1,it}, \dots, \xi_{Q,it}]'$ and assume

$$\mathbf{v}_{it} \sim \mathcal{N}_{J+Q}(\mathbf{0}, \mathbf{\Sigma}). \quad (26)$$

For the one-sided inefficiency terms we assume:

$$u_{1,it} \sim \mathcal{N}^+(0, \sigma_u^2), \eta_{q,it} \sim \mathcal{N}^+(0, \omega_q^2), q = 1, \dots, Q, \quad (27)$$

distributed independently of each other, each truncated at 0 from below. We assume them to be independent of all other random variables and the regressors. Since there are cross-equational restrictions in terms of the parameters in the IDF and the FOCs in (7) and (24), one cannot estimate these equations separately using standard stochastic cost frontier models (because the one-sided inefficiency terms in these models are non-negative) although the error components (both inefficiency and noises) are assumed to be independent of each other. There are $(J + Q)$ endogenous variables (X^g and Z) with $(J + Q)$ equations consisting of (7), (24), and (8). This system of equations differs from the one in Model 1 which does not include the $(J - 1)$ FOCs, and assumes input ratios \tilde{X}^g to be exogenous. If this assumption does not hold, estimates from Model 1 will be inconsistent.

Similar distributional assumptions are made to estimate the system in Models 2a and 2 as well as Models 0a and 0.

5.1 Implementation of MCMC

In this subsection we provide a brief discussion about the implementation of the MCMC technique for the Bayesian analysis of our models. Instead of tailoring our discussion for each model, we focus on the BP approach, viz., Model 1a. Since the model in (7) is linear in parameters, to simplify our discussion, we denote all the parameters by $\boldsymbol{\beta}$ and the corresponding regressors by \mathbf{X} which includes the linear, squares and cross-products of good outputs (Y), ratios of good inputs (\tilde{X}^g), bad inputs (X^b) and time. The same parameters appear in the FOCs in (24). We use the same procedure and denote all the parameters in (8) by $\boldsymbol{\delta}$ and the corresponding regressors by \mathbf{Z} which includes the linear, squares and cross-products of good outputs, bad inputs and time.

Using the above simplification, we write the IDF and FOCs associated with Model 1a in (7), (24) and (8) as:

$$\begin{aligned} x_{1,it} &= \mathbf{X}'_{it}\boldsymbol{\beta} + v_{1,it} + u_{1,it} \\ \tilde{\mathbf{w}}_{it} &= h(\mathbf{X}_{it}, \boldsymbol{\beta}) + \tilde{\mathbf{v}}_{it} \\ \mathbf{z}_{it} &= \mathbf{Z}'_{it}\boldsymbol{\delta} + \boldsymbol{\xi}_{it} + \boldsymbol{\eta}_{it} \end{aligned} \quad (28)$$

where $\tilde{\mathbf{v}}_{it} = (v_{2,it}, \dots, v_{J,it})'$, \mathbf{X}_{it} denotes all regressors of the IDF, $\tilde{\mathbf{w}}$ denotes the ratio of input prices (in log), \mathbf{Z}_{it} denotes the regressors and the column of ones in (8) and $\boldsymbol{\beta}, \boldsymbol{\delta}$ are parameter vectors. We denote the data by \mathcal{D} and $\mathbf{\Omega} = \text{diag}[\omega_q^2, q = 1, \dots, Q]$, $\omega_q^2 =$ pre-truncated var of $\eta_{q,it}$, $q = 1, \dots, Q$. Implementation of MCMC relies on drawing from the following posterior conditional distributions:

- (i) $\boldsymbol{\beta} | \tilde{\mathbf{w}}, u_1, \boldsymbol{\eta}, \sigma_u, \mathbf{\Sigma}, \mathbf{\Omega}, \mathcal{D}$,
- (ii) $\boldsymbol{\delta} | \boldsymbol{\beta}, \tilde{\mathbf{w}}, u_1, \boldsymbol{\eta}, \sigma_u, \mathbf{\Sigma}, \mathbf{\Omega}, \mathcal{D}$,
- (iii) $\mathbf{\Sigma} | \boldsymbol{\beta}, \boldsymbol{\delta}, \tilde{\mathbf{w}}, u_1, \boldsymbol{\eta}, \sigma_u, \mathbf{\Omega}, \mathcal{D}$,

- (iv) $\Omega|\beta, \delta, \tilde{w}, u_1, \eta, \sigma_u, \Sigma, \mathcal{D}$,
- (v) $\sigma_u|\beta, \delta, \tilde{w}, u_1, \eta, \sigma_u, \Sigma, \Omega, \mathcal{D}$,
- (vi) $u_1|\beta, \delta, \tilde{w}, \eta, \sigma_u, \Sigma, \Omega, \mathcal{D}$,
- (vii) $\eta|\beta, \delta, \tilde{w}, u_1, \sigma_u, \Sigma, \Omega, \mathcal{D}$.
- (viii) $\tilde{w}|\beta, \delta, \tilde{w}, u_1, \eta, \sigma_u, \Sigma, \Omega, \mathcal{D}$.

Models that do not include the FOCs do not rely on \tilde{w} and, therefore, we can omit step (viii). Details are provided in the Appendix.

Imposition of the monotonicity constraints

Monotonicity constraints in (11) - (15) can be written compactly as

$$\mathbf{c}_{it}(\mathcal{D}_{it}, \beta, \delta) \leq \mathbf{0}, \quad (29)$$

where \mathbf{c}_{it} is a vector function in $\mathbb{R}^{J-1+Q(M+K)}$ and $\mathcal{D}_{it} = [\tilde{\mathbf{x}}'_{it}, \mathbf{y}'_{it}, \mathbf{x}^b'_{it}, \mathbf{z}'_{it}]$ denotes a particular observation. The total number of restrictions is $N(J-1+Q(M+K))$ where $N = nT$, n being the number of cross-sectional units and T is the number of times each cross-sectional unit is observed.

Suppose the mean of the data is $\bar{\mathcal{D}} = \mathbf{0}$ for simplicity. First the restrictions in (29) are imposed at the mean, viz., $\mathbf{c}_{it}(\mathbf{0}, \beta, \delta) \leq \mathbf{0}$ and then at points $\pm \mathbf{r}$, where \mathbf{r} is a vector which is conformable with \mathcal{D}_{it} whose typical element, say r_p , extends from the minimum to the maximum of the p th element of \mathcal{D}_{it} ($p = 1, \dots, J-1+Q+M+K$). There is a number of \mathcal{N} such points and \mathcal{N} is selected so that a simple acceptance algorithm to enforce (29) does not take more than 10,000 rejections per observation. We have found that $\mathcal{N} = 50$ such points are enough and the average rejection rate (per observation and MCMC iteration) is about 1,500 when the Girolami and Calderhead (2012, GC) update is used.¹²

Priors

Our prior of β is specified as:

$$\beta \sim \mathcal{N}_K(\bar{\beta}, \mathbf{V}_\beta) \quad (30)$$

subject to the restriction that it is defined over the region defined by the monotonicity constraints in (11) -(15). We choose $\bar{\beta} = \mathbf{0}$ and $\mathbf{V}_\beta = 10^3 \mathbf{I}$ so that these priors are minimally informative. Similarly, the prior for δ is assumed to be flat.

For Σ we assume that $\Sigma = \begin{bmatrix} \sigma_{v_1}^2 & \mathbf{0}' \\ \mathbf{0} & \tilde{\Sigma} \end{bmatrix}$. For all variance parameters we assume a prior in the inverted gamma family:

$$\frac{Q_d}{\sigma_d^2} \sim \chi^2(\underline{N}_d) \quad (31)$$

for $d \in \{v, u, \lambda, \lambda, \varepsilon_j, \tilde{w}\}$. For $\tilde{\Sigma}$ we assume a prior in the inverted Wishart family: $p(\tilde{\Sigma}) \propto |\tilde{\Sigma}|^{(\underline{N}+J)/2} \exp\left\{-\frac{1}{2}\text{tr}\underline{\mathbf{A}}\tilde{\Sigma}^{-1}\right\}$. As a baseline prior we choose $\underline{N}_d = \underline{N} = 1$ and $\underline{Q}_d = 0.001$. Moreover $\underline{\mathbf{A}} = 0.001\mathbf{I}$, which is compatible with the existence of the posterior and is minimally informative. **Mike, Are these .001 or 10^3 ?**

¹²Without the GC update we need $\mathcal{N} = 250$ points and the average rejection rate (per observation and MCMC iteration) is about 8,500. Details on the numerical performance of the algorithm, relative numerical efficiency and convergence diagnosis are available upon request.

5.2 Estimation of Efficiency

To estimate inefficiency we examine the conditional distribution of \mathbf{u}_1 which is truncated normal:

$$u_{1,it}|\text{data,parameters} \sim N^+ \left(\frac{x_{1,it} - \mathbf{X}'_{it}\boldsymbol{\beta}}{\sigma_u} \sigma_v^2, \frac{\sigma_v^2 \sigma_u^2}{\sigma_v^2 + \sigma_u^2} \right). \quad (32)$$

This conditional distribution is used explicitly in MCMC to generate draws for $u_{1,it}$. Given MCMC draws $\{u_{1,it}^{(s)}, s = 1, \dots, S\}$, efficiency draws are generated as $r_{it}^{(s)} = \exp(-u_{1,it}^{(s)})$, $s = 1, \dots, S$. The posterior mean of technical efficiency across all draws will be:

$$r_{it}^\eta = S^{-1} \sum_{s=1}^S \exp(-u_{1,it}^{(s)}). \quad (33)$$

To estimate environmental efficiency for each bad output Z_q , we consider the conditional distribution of $\eta_{q,it}$, $q = 1, \dots, Q$ which is truncated normal:

$$\eta_{q,it}|\text{data,parameters} \sim N^+ \left(\frac{z_{q,it} - \mathbf{Z}'_{it}\boldsymbol{\delta}_q}{\sigma_{\xi,q}} \sigma_{\eta,q}^2, \frac{\sigma_{\xi,q}^2 \sigma_{\eta,q}^2}{\sigma_{\xi,q}^2 + \sigma_{\eta,q}^2} \right). \quad (34)$$

This conditional distribution is used explicitly in MCMC when we draw from it to generate draws for $\eta_{q,it}$. Suppose the MCMC draws are $\{\eta_{q,it}^{(s)}, s = 1, \dots, S\}$. Then the posterior mean of environmental efficiency associated with z_q , $s = 1, \dots, S$ across all draws will be:

$$E_{q,it} = S^{-1} \sum_{s=1}^S \exp(-\eta_{q,it}^{(s)}) \quad (35)$$

which can be taken as a predictor of environmental efficiency for bad output q and observation corresponding to (i, t) . A kernel density of $E_{q,it}$ across observations for all $q = 1, \dots, Q$ then provides the sample distribution of the predictor for environmental efficiency.

6 Empirical results

Our primary focus in this paper is to estimate technical and environmental efficiency. Note that Models 0 and 0a cannot capture environmental inefficiency. Since environmental efficiency is an important issue in models with bad outputs, the use of a single equation representation of the technology in Models 0 and 0a might not be very appealing. In contrast Models 1 and 1a not only distinguish between technical and environmental efficiency, we can also estimate environmental efficiency for each bad output. On the other hand, Models 2 and 2a separate technical and environmental efficiency but both of them are radial (i.e., technical inefficiency overuses all good inputs by the same proportion and environmental inefficiency leads to overproduction of all bad outputs by the same proportion).

In addition to estimating technical and environmental efficiency we also estimate and report estimates of returns to scale (RTS) and technical change (TC). These are important features of any production technology and are of general interest to all production economists. We compare these measures as well as technical and environmental efficiency across the models. Finally, we do a Bayesian model selection test to check which model better fits the data.

RTS in an IDF is computed from $[\sum_m \partial x_1^q / \partial y_m]^{-1}$ in all the models. Estimates of observation-specific RTS tell us whether the scale size of a plant is optimal (RTS=1) or not. If RTS exceeds unity the plant can benefit from expansion. The opposite is true when RTS is less than unity. Similarly, TC defined as $TC = \partial x_1 / \partial t$ which

(similar to a cost function model) is a measure of cost diminution over time, holding everything else unchanged. A negative (positive) value of TC would thus mean technical progress (regress).

Summary statistics of these results are presented in Table 3. Finally, kernel densities of RTS, TC and TE are reported for all the models in Figures 1-4.

First, we examine RTS estimates from each model. Judging by the mean values of RTS (reported in Table 3), it is clear that RTS from Models 0 and 0a are too small (the mean being close to 0.3). Correcting for endogeneity of inputs did not change the RTS results much. These can be clearly seen from the kernel density plots of RTS in panels a and b of Figure 1. The maximum value of RTS is less than 0.5 which is clearly unrealistic. A dramatic change in the estimated values of RTS is observed when endogeneity is corrected and a BP approach is used (Model 1a). The RTS values are more realistic in the sense that some are close to unity (which gives the optimal scale size). Estimates of RTS from Models 2 and 2a are much lower than unity. The spread of RTS after the endogeneity correction (Model 2a) is quite large (the range is from almost 0 to 1.2).

Next we examine TC. The mean value of TC is quite small in every model (less than 1% per annum). However, a close look at the density plots show large variations in the estimates of TC within each model. Unlike RTS, variations in TC across models are quite small. A substantial reduction in the variability of TC is observed after endogeneity correction in Models 0 and 1 (0a and 1a) but not in Model 2a. Technical progress is observed for more than 50% of the observations in all three models after the endogeneity correction.

Table 3: Results for various models

Model	RTS		TC		TE		EE (SO_2)		EE (NO_x)	
	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
0	0.323	0.076	-0.0005	0.0232	0.871	0.0157	—	—	—	—
1	0.642	0.061	0.0037	0.0222	0.955	0.0068	0.910 ^a	0.023	0.891 ^b	0.0300
2	0.512	0.132	0.0037	0.0224	0.913	0.0135	0.922 ^c	0.022	—	—
0a	0.303	0.079	-0.0091	0.0151	0.822	0.0144	—	—	—	—
1a	0.957	0.052	-0.0080	0.0087	0.915	0.0230	0.968 ^a	0.0068	0.973 ^b	0.0108
2a	0.539	0.221	-0.0140	0.0232	0.891	0.0128	0.905 ^c	0.0157	—	—

1. The table reports sample means and sample standard deviations. RTS is returns to scale, TC is technical change, TE is technical efficiency and EE is environmental efficiency.

2. (a) corresponds to SO_2 , (b) corresponds to NO_x and (c) corresponds to overall environmental efficiency.

Estimates of summary statistics of TE are reported in Table 3. In general the mean values of TE are quite high (80% or higher) across models. Predicted TE values are the highest in Models 1 and 1a. A close look at Figure 3 shows that the distributions of TE are quite tight although they differ substantially across models. Note that the TE is input-oriented which means that input-usage, on average, is 91.5% efficient in Model 1a. Alternatively, Model 1a predicts that cost is increased, on average, by 8.5% due to input over-use.

Finally, we report environmental efficiency (EE) for both SO_2 and NO_x in Figure 4. Note that Model 0 cannot separate TE from EE which might be the reason why TE in Model 0 is the lowest among the three models. EE in Models 2 and 2a is radial thereby meaning that EE in the production of SO_2 and NO_x are identical by construction. In both models estimated values of EE are quite high (above 90% on average). There is some scope for improvement (reduction in SO_2 and NO_x), especially for some plants at the lower tails of the distributions. Models 1 and 1a give separate estimates of EE for SO_2 and NO_x . Again these estimates are quite high (90% or more, on average). Based on these findings, we conclude that the plants are quite efficient from both technical and environmental points of view.

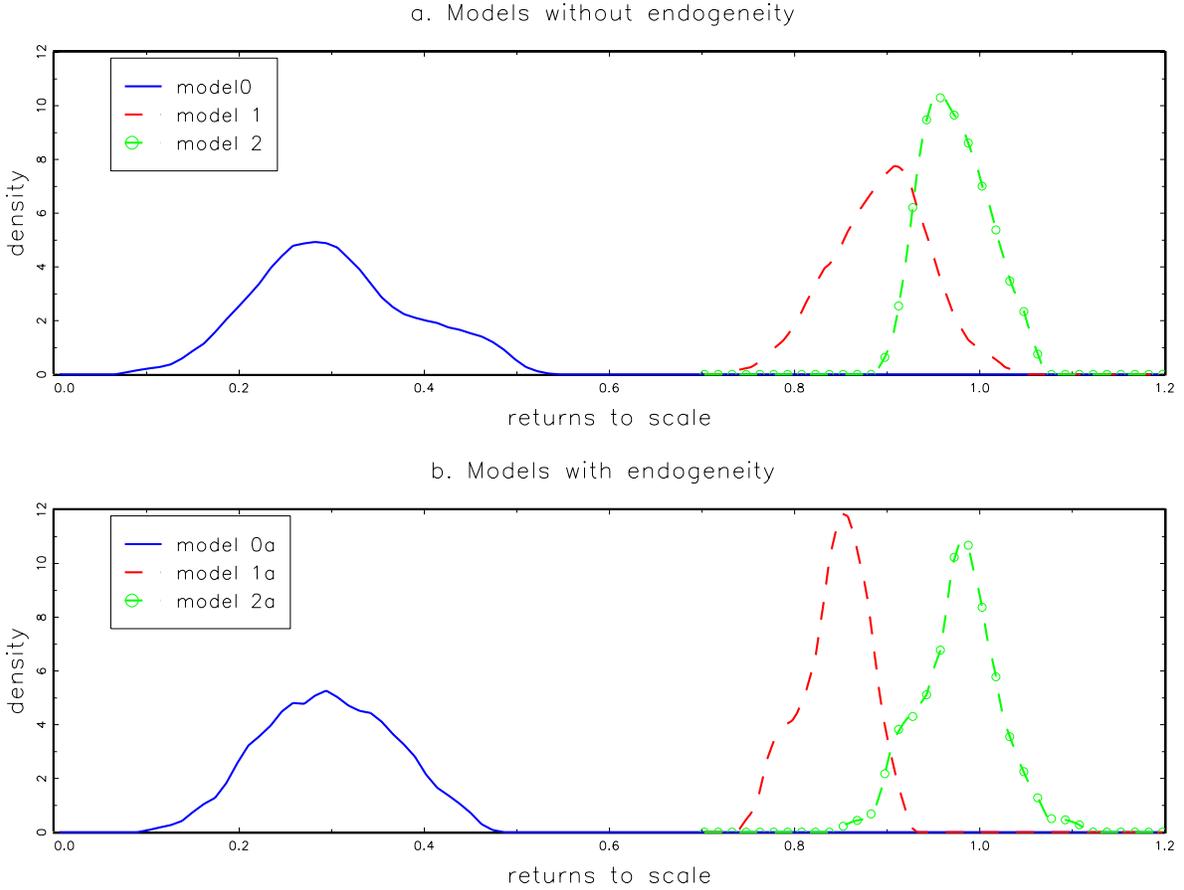


Figure 1: Returns to Scale

6.1 Model Comparison

Since we used three different models it is natural to ask how these models perform in cross-validating predictive exercises. The comparison can be made in terms of fit using Bayes factors which is compatible with forecasting or predictive type exercises. Suppose $\mathcal{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)$ denotes a data set for firms $1, \dots, n$. The marginal likelihood for a given model $m \in \{0, 1, 2, \dots, M\}$ is

$$\mathcal{M}_m(\mathcal{Y}) = \int L_m(\boldsymbol{\theta}_m; \mathcal{Y}) p_m(\boldsymbol{\theta}_m) d\boldsymbol{\theta}, \quad (36)$$

where $L_m(\cdot)$ is the likelihood function for model m and $p_m(\cdot)$ is the prior probability of the parameters in model m . The Bayes factor in favor of model $m \in \{1, \dots, M\}$ against model 0 is given by:

$$BF_{m:0} = \frac{\mathcal{M}_m(\mathcal{Y})}{\mathcal{M}_0(\mathcal{Y})}, \quad m = 1, \dots, M.$$

The posterior probability in favor of any given model is given by:

$$P_m(\mathcal{Y}) = \frac{\mathcal{M}_m(\mathcal{Y})}{\sum_{m'=0}^M \mathcal{M}_{m'}(\mathcal{Y})}, \quad m = 0, 1, 2, \dots, M \quad (37)$$

which summarizes the evidence in the data in favor of a specific model m . To use this we need to compute the marginal likelihoods in (36). If the posterior distribution is denoted by $p(\boldsymbol{\theta}|\mathcal{Y})$ then we have the so-called

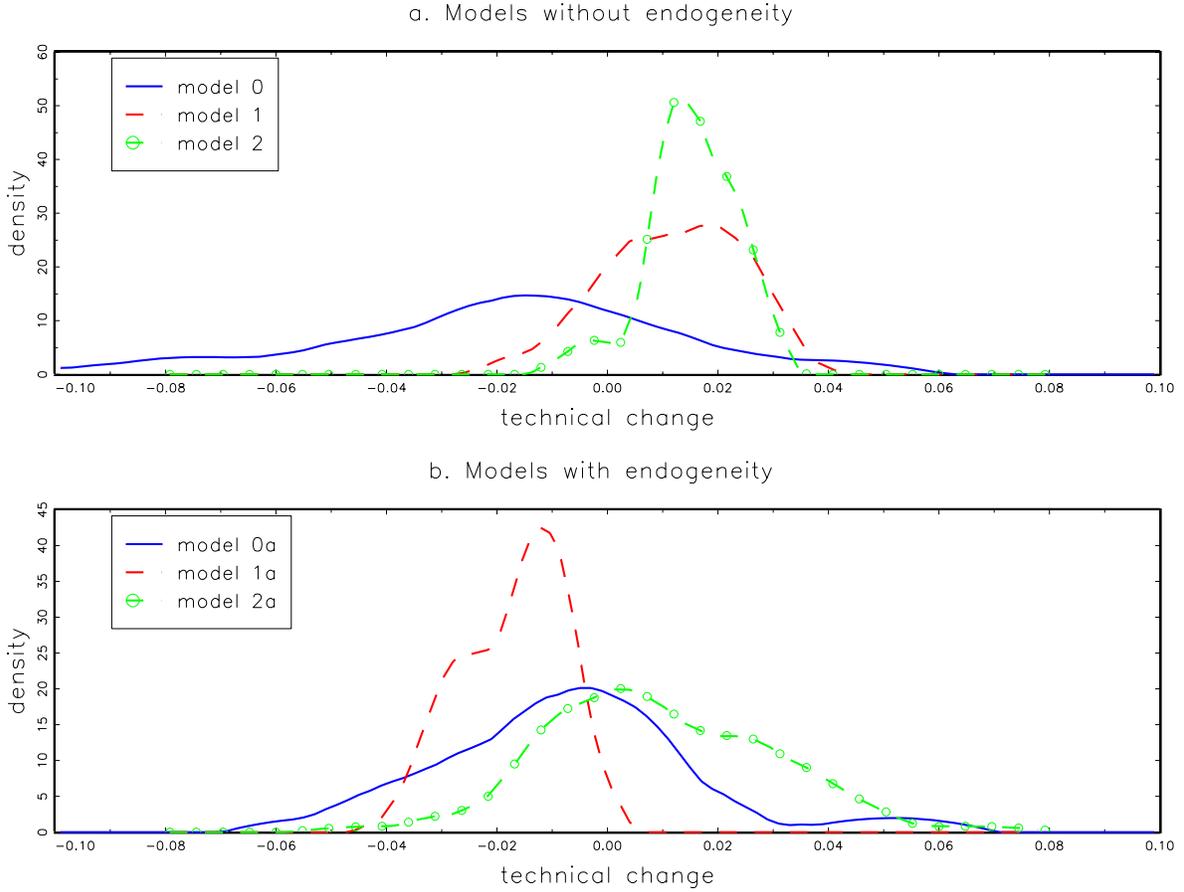


Figure 2: Technical Change

candidate's formula:

$$\mathcal{M}_m(\mathcal{Y}) = \frac{L_m(\boldsymbol{\theta}_m; \mathcal{Y}) p(\boldsymbol{\theta}_m)}{p_m(\boldsymbol{\theta} | \mathcal{Y})} \quad (38)$$

which holds identically for all $\boldsymbol{\theta}_m$. Therefore, we can fix $\boldsymbol{\theta}_m = \bar{\boldsymbol{\theta}}_m$ where $\bar{\boldsymbol{\theta}}_m$ is any point of high posterior mass, for example the posterior mean. Then computation of (38) is, in principle, straightforward except for the fact that the denominator is unknown. Since all posterior conditional distributions are in closed form, as we show in the Technical Appendix, Chib's (1994) approach can be used by running some additional MCMC chains. The results of model comparison are shown in Table 4. Since the probability in favor of Model 1a is 0.985, it is safe to say that it fits the data best, followed by model 2a (whose posterior probability is 1.4%).

Table 4: Posterior model probabilities

Model	$P_m(\mathcal{Y})$
0	0.000
1	0.000
2	0.000
0a	0.010
1a	0.985
2a	0.014

There are two technical questions about these posterior model probabilities. First, how sensitive are they

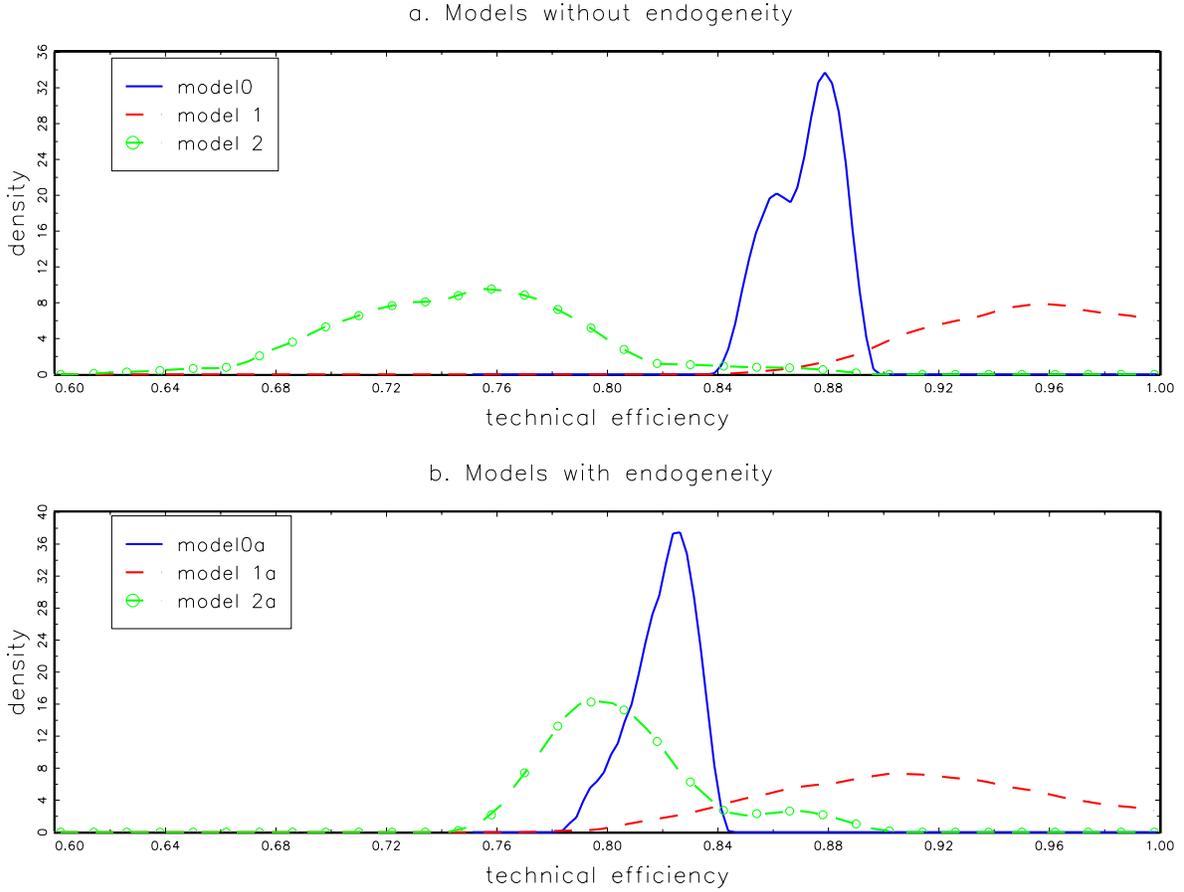


Figure 3: Technical Efficiency

with respect to the prior? Second, how can the models be compared since they, for example, Models 0 and 1 or Models 0a and 1a, have different endogenous variables? The first question is resolved by varying the parameters of the prior 10,000 times thus generating 10,000 different priors.¹³ Approximate posteriors corresponding to these alternative priors are computed using the Sampling-Importance-Resampling (SIR) algorithm of Rubin (1988). Bayes factors and posterior model probabilities can then be computed using (38). Our results indicate that the 99% confidence interval for the posterior model probability for Model 1a is 92% to 100%. From that point of view the results reported here are impressively robust.

As for the second point, let the probability distribution of the data for Model 0 is $p_0(\mathbf{X}|\mathbf{Y}, \tilde{\mathbf{X}}, \mathbf{Z}, \mathbf{X}^b)$ and for Model 1 it is $p_1(\mathbf{X}, \mathbf{Z}|\mathbf{Y}, \tilde{\mathbf{X}}, \mathbf{X}^b)$. Since the distributions are different we cannot compare the models in this form. Instead we have to compute $\tilde{p}_1(\mathbf{X}|\mathbf{Y}, \mathbf{Z}, \tilde{\mathbf{X}}, \mathbf{X}^b)$ from Model 1 and compare it to $p_0(\mathbf{X}|\mathbf{Y}, \tilde{\mathbf{X}}, \mathbf{Z}, \mathbf{X}^b)$ where $\tilde{p}_1(\mathbf{X}|\mathbf{Y}, \mathbf{Z}, \tilde{\mathbf{X}}, \mathbf{X}^b) = \frac{p_1(\mathbf{X}, \mathbf{Z}|\mathbf{Y}, \tilde{\mathbf{X}}, \mathbf{X}^b)}{p_z(\mathbf{Z}|\mathbf{Y}, \tilde{\mathbf{X}}, \mathbf{X}^b)}$. The computation of $\tilde{p}_1(\cdot)$ relies on, (i) the numerator, which is easy to evaluate from the joint distribution of the data in Model 1, and (ii) the denominator which can be computed by integrating the joint distribution with respect to \mathbf{X} . Due to independence, this integration reduces to computing a $(J - 1)$ -dimensional integral for each observation. In our case this is a univariate integral which can be evaluated using the standard quadrature procedure.¹⁴

¹³We vary the parameters $\underline{Q}, \underline{N}$ uniformly in the interval 10^{-5} to 100 and 0.01 to 100 respectively.

¹⁴We used a 40-point Gauss-Kronrod rule as implemented in IMSL.

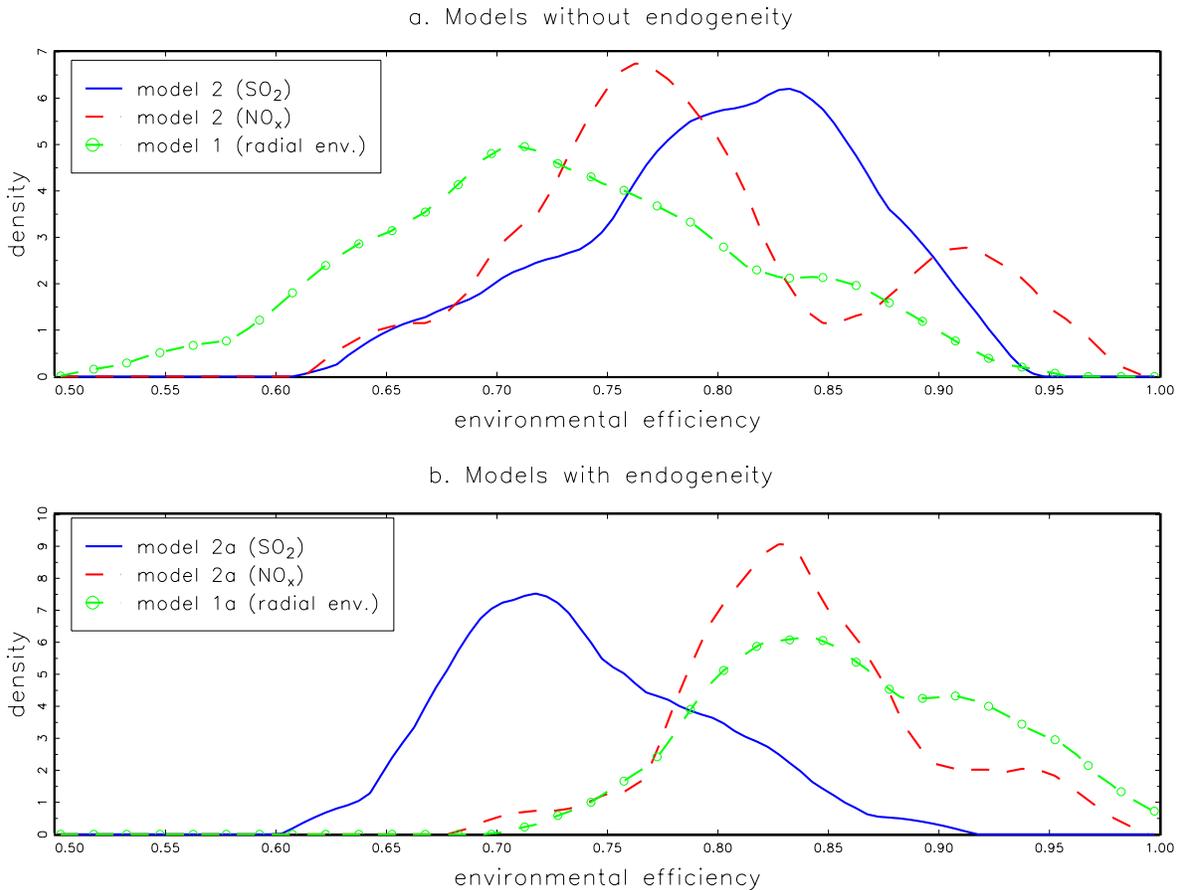


Figure 4: Environmental Efficiency

7 Conclusion

In this paper we consider an environmental production process (electricity generation) in which firms intend to produce electricity (desired/good outputs) but the production process is such that it automatically produces undesirable/unintended outputs such as emission pollutants (SO_2 and NO_x). We separate the technology for the production of the good output from the production of undesirable outputs. The former describes how good inputs are transformed into desirable outputs. The latter can be viewed as a by-production (BP) technology. Like stochastic frontier functions, the BP technology specifies that there is a certain minimal amount of the undesirable by-product that is produced, given the quantities of certain good outputs and inputs. The presence of (environmental) inefficiency in the BP technology would thus generate more than this minimal amount of the undesirable output. Similarly, the presence of technical inefficiency may imply that, given inputs, less than the maximal possible amount of good output is produced. Alternatively it may imply that more than the minimal amount of inputs are used to produce a given level of desired output. We used the BP approach, which is a composition of desired and undesired production technologies, and estimate both technical and environmental efficiencies. Given that electricity, the good output in our application, is demand determined, we treat it as exogenous and use an input distance function to represent the electricity generation technology. We use a series of models and examine their appropriateness in capturing the features of environmental production technology. First, we consider the modeling approach in which production of both good and bad outputs comes from the same technology. This is followed by a model which separates the production of good outputs from the bad outputs by specifying two separate technologies (the BP approach). Finally, we extend the BP approach further

in which the production of each bad output is allowed to have its own production process. The last model allows us to estimate technical efficiency and environmental efficiency (for each bad output). On the other hand, the first model could not separate technical efficiency from environmental efficiency.

We also address the endogeneity of inputs by adding the first-order conditions of cost minimization (which is standard in the electricity generation literature going back to Nerlove (1965)). We did this in each of the above-mentioned models and estimate each of them using a system approach. More specifically, we use efficient Bayesian MCMC technique to estimate the technologies and both types of inefficiencies using a system approach that takes endogeneity into account. Finally, we did a model comparison exercise and found that the BP approach with separate technology for each bad output fits the data best.

Average technical efficiency (TE) is found to range from 82.2% to 95.5%. Barring a few outliers, distributions of TE are found to be quite concentrated, with noticeable differences across different models with and without the endogeneity correction. Environmental efficiency estimates based on the BP approach (without the endogeneity correction) tend to vary within models but across model variations are found to be quite small. We found the distribution of SO_2 centered around 96.8%, while the distribution of NO_x centered around 97.3% in the model that allowed two separate technologies for SO_2 and NO_x and corrected for endogeneity of inputs. Based on these findings, we conclude that the electricity generating plants operated quite efficiently during the period of our study. That is, the scope for reducing SO_2 and NO_x , holding everything else unchanged, was limited.

In addition to TE and EE, we also compute returns to scale (RTS) and technical change (TC). RTS estimates differ widely and some models produce estimates that are quite low. The mean value of RTS in our preferred model is found to be 0.957 with a standard deviation of 0.052. Estimates of TC are found to center around zero with noticeable variations within each model. About 50% of the plants are found to have technical progress (the 2 standard deviation range in the preferred model gives a range of 2.5% to -0.9%). Variations of TC across models are found to be quite small and are reduced after the endogeneity correction.

References

- [1] Agee, Atkinson, Crocker and Williams (2014), Non-separable pollution control: Implications for a CO_2 emissions cap and trade system, *Resource and Energy Economics* 36, pp. 64-82.
- [2] Atkinson and Dorfman (2005), Bayesian measurement of productivity and efficiency in the presence of undesirable outputs: crediting electric utilities for reducing air pollution, *Journal of Econometrics* 126, pp. 445-468.
- [3] Caves, D.W., Christensen, L.R. and J. A. Swanson (1981), Productivity Growth, Scale Economies, and Capacity Utilization in U.S. Railroads, 1955-74, *American Economic Review* 71, pp. 994-1002.
- [4] Chib, S. (1995), Marginal Likelihood from the Gibbs Output, *Journal of the American Statistical Association* 90, pp. 1313-1321.
- [5] Chung, Y.H., Färe, R., Grosskopf, S. (1997), Productivity and undesirable outputs: a directional distance function approach. *Journal of Environmental Management* 51, pp. 229-240.
- [6] Färe, R., Grosskopf, S. (1998), Shadow pricing of good and bad commodities, *American Journal of Agricultural Economics* 80, pp. 584-590.
- [7] Färe, R., Grosskopf, S. (2003), Non-parametric Productivity Analysis with Undesirable Outputs: Comment, *American Journal of Agricultural Economics* 85, pp. 1070-1074.
- [8] Färe, R., Grosskopf, S., Noh, D.-W., Weber, W.L. (2005), Characteristics of a polluting technology: theory and practice. *Journal of Econometrics* 126, pp. 469-492.

- [9] Färe, R., S. Grosskopf, and D. Margaritis (2008), Productivity and efficiency: Malmquist and more, in H. Fried, C. A. K. Lovell, and S. Schmidt, eds., *The Measurement of Productive Efficiency*, chapter 5, pp. 522-621, Oxford: Oxford University Press, 2nd edition.
- [10] Fernández, G. Koop and M.F.J. Steel (2002), Multiple-Output production with undesirable outputs: an application to nitrogen surplus in agriculture, *Journal of the American Statistical Association* 97, p. 432-442.
- [11] Fernández C., G. Koop and M.F.J. Steel (2005), Alternative Efficiency Measures for Multiple-Output Production, *Journal of Econometrics* 126, pp. 411-444.
- [12] Førsund, F. (2009), Good modelling of bad outputs: pollution and multiple-output production, *International Review of Environmental and Resource Economics* 3, pp. 1-38.
- [13] Geweke, J. (1992), Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In *Bayesian Statistics 4*, (ed. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith). Clarendon Press, Oxford, UK.
- [14] Girolami, M. and B. Calderhead (2011), Riemann manifold Langevin and Hamiltonian Monte Carlo methods, *Journal of the Royal Statistical Society B*, 73, pp. 123-214.
- [15] Hailu, A., Veeman, T.S. (2000), Environmentally Sensitive Productivity Analysis of the Canadian Pulp and Paper Industry, 1959-1994: An Input Distance Function Approach. *Journal of Environmental Economics and Management* 40, pp. 251-274.
- [16] Kumbhakar, S.C. (2012), Specification and Estimation of Primal Production Models, *European Journal of Operational Research* 217, pp. 509-518.
- [17] Kumbhakar, S.C. and Lovell, C.A.K. (2000), *Stochastic Frontier Analysis*, Cambridge University Press, New York.
- [18] Lee, J.-D., Park, J.-B., Kim, T.-Y. (2002), Estimation of the shadow prices of pollutants with production/environment inefficiency taken into account: a nonparametric directional distance function approach, *Journal of Environmental Management* 63, pp. 365-375.
- [19] Murty, S., Russell, R.R., Levkoff, S.B. (2012), On modeling pollution-generating technologies. *Journal of Environmental Economics and Management* 64, pp. 117-135.
- [20] Nerlove, M. (1965), *Estimation and Identification of Cobb-Douglas Production Functions*, Rand McNally Company, Chicago, Illinois.
- [21] Pasurka, C.A. (2006), Decomposing electric power plant emissions within a joint production framework, *Energy Economics* 28, pp. 26-43.
- [22] Reinhard, S., Lovell, C.A.K, Thijssen, G. (1999), Econometric estimation of technical and environmental efficiency: an application to Dutch dairy farms. *American Journal of Agricultural Economics* 81, pp. 44-60.
- [23] Reinhard, S., Lovell, C.A.K, Thijssen, G. (2000), Environmental efficiency with multiple environmentally detrimental variables; estimated with SFA and DEA. *European Journal of Operational Research* 121, pp. 287-303.
- [24] Simar, L., Vanhems, A., and Wilson, P.W. (2012), Statistical inference for DEA estimators of directional distances, *European Journal of Operational Research*, 220, pp. 853-864,
- [25] Tsionas, E. G., Malikov, E., and Kumbhakar, S. C. (2014). Productivity index based on the directional output distance function with undesirable outputs: Direction, normalization and the Jacobian. Working Paper, Binghamton University.
- [26] Vardanyan, M., Noh, D.-W. (2006), Approximating pollution abatement costs via alternative specifications of a multioutput production technology: a case of the US electric utility industry. *Journal of Environmental Management* 80, pp. 177-190.

[27] Verdinelli, I. and Wasserman, L. (1995), Computing Bayes factors using a generalization of the Savage-Dickey density ratio, *Journal of the American Statistical Association* 90, pp. 614-618.

Supplementary Materials for Online Publication

Technical Appendix

In this Appendix we provide details on how to draw from the conditional posterior distributions in (i) through (viii). Drawing from the conditional posterior distributions results in a Gibbs sampling scheme which, under very general conditions, provides samples that converge in distribution to the joint posterior distribution of all structural parameters and latent variables.

(i) Drawing from the posterior conditional of β

To construct a proposal for β the first equation is estimated by GMM using as instruments all exogenous variables, their squares and cross-products. Denote the GMM estimator by $\hat{\beta}_{GMM}$. Generally this is a poor approximation to the posterior due to the lack of ignoring the second set of equations. After linearizing the second set of equations in (28) with respect to β we end up with a system of the form:

$$\begin{aligned} x_{1,it} &= \mathbf{X}'_{it}\beta + v_{1,it} + u_{1,it} \\ \mathbf{a}_{it} &= \mathbf{F}'_{it}\beta + \tilde{v}_{it} \end{aligned}, \quad (39)$$

where $\mathbf{X}'_{it} = [\tilde{\mathbf{x}}'_{it}, \tilde{\mathbf{z}}'_{it}]$ so that the first elements are logs of deviations of good inputs from the first good input. Here, \mathbf{a}_{it} and \mathbf{F}_{it} correspond to terms related to a Taylor series expansion of the second equation in (23), that is $\mathbf{a}_{it} = \tilde{\mathbf{w}}_{it} - h(\bar{\mathbf{X}}_{it}, \hat{\beta}_{GMM})$ and $\mathbf{F}_{it} = \nabla h(\bar{\mathbf{X}}_{it}, \hat{\beta}_{GMM})$, where $\bar{\mathbf{X}}$ is the point of reference for the linearization. The formulation in the second set of equations of (39) allows for the presence of equations for bad outputs provided it is augmented to include one-sided error terms $\eta_{q,it}$. In this section we are interested in crafting a good proposal density for β so the inclusion of equations for bad outputs is subsumed in the second equation of (39).

We write the system above in the following form¹⁵:

$$\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_{Q-1} \end{bmatrix} = \begin{bmatrix} \mathbf{X} \\ \mathbf{F}_1 \\ \vdots \\ \mathbf{F}_{Q-1} \end{bmatrix} \beta + \begin{bmatrix} \mathbf{v}_1 \\ \tilde{\mathbf{v}}_1 \\ \vdots \\ \tilde{\mathbf{v}}_{Q-1} \end{bmatrix} + \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{O} \\ \vdots \\ \mathbf{O} \end{bmatrix} \quad (40)$$

or in compact form as:¹⁶

$$\mathbb{Y} = \mathbb{X}\beta + \mathbb{V} \quad (41)$$

with $E(\mathbb{V}\mathbb{V}') = \Sigma \otimes \mathbf{I}_{nT}$. From standard results about the multivariate regression model (Zellner, 1971, pp. 241-243) we have:

$$\beta | \Sigma, \mathbb{Y} \sim \mathcal{N}_K \left(\left[\mathbb{X}' (\Sigma^{-1} \otimes \mathbf{I}_{nT})^{-1} \mathbb{X} + \mathbf{V}_{\beta}^{-1} \right]^{-1} \left[\mathbb{X}' (\Sigma^{-1} \otimes \mathbf{I}_{nT})^{-1} \mathbb{Y} + \mathbf{V}_{\beta}^{-1} \bar{\beta} \right], \mathbb{X}' (\Sigma^{-1} \otimes \mathbf{I}_{nT})^{-1} \mathbb{X} \right), \quad (42)$$

¹⁵Columns of $\mathbf{F}_1, \dots, \mathbf{F}_{Q-1}$ are padded with zeros in the appropriate places so that these matrices have the same dimensionality with \mathbf{X} and are conformable with β .

¹⁶Note that we consider $\mathbf{x}_1 - \mathbf{u}_1$ together in defining \mathbb{Y} .

where the prior of β is specified as:

$$\beta \sim \mathcal{N}_K(\bar{\beta}, \mathbf{V}_\beta). \quad (43)$$

We choose $\bar{\beta} = \mathbf{O}_K$ and $\mathbf{V}_\beta = 0.001\mathbf{I}$ so that these priors are minimally informative. The posterior conditional distribution of the covariance matrix is in the inverted Wishart family:

$$p(\Sigma^{-1}|\beta, \mathbb{Y}) \propto |\Sigma^{-1}|^{nT - \frac{K+1}{2} + N} \exp\left\{-\frac{1}{2}(\mathbf{A} + \underline{\mathbf{A}})\Sigma^{-1}\right\}, \quad (44)$$

where $\mathbf{A} = \mathbb{V}(\beta)\mathbb{V}(\beta)'$, $\mathbb{V}(\beta) = \mathbb{Y} - \mathbb{X}\beta$. Using the conditional posterior distributions for β and Σ we implement Gibbs sampling which, however, is only an approximate in this case due to the linearization of the FOCs.

We introduce the following modification of this basic algorithm. Instead of centering the proposal around GMM and then using posterior moments to configure a proposal we use a Girolami and Calderhead (2012, GC) algorithm. The GC algorithm starts at the GMM estimator and the MCMC runs until convergence. Depending on the model and the subsample this takes 20,000 to 100,000 iterations. Then we run one million MCMC iterations to obtain final results for posterior moments and densities of parameters and functions of interest. We find that the GC algorithm performs vastly superior relative to the standard Metropolis-Hastings algorithm and autocorrelations are much smaller.

(ii) Drawing from the conditional posterior of δ

Drawing from this posterior conditional relies on the system of equations: $z_{it} = \mathcal{Z}'_{it}\delta + \xi_{it} + \eta_{it}$. Conditional on the data and η we have:

$$\delta|\eta, \Omega, \mathcal{D} \sim N(\hat{\delta}, \mathbf{V}_\delta), \quad (45)$$

where $\hat{\delta} = (\mathcal{Z}'\Omega^{-1}\mathcal{Z})^{-1}\mathcal{Z}'\Omega^{-1}(z - \eta)$, $\mathbf{V}_\delta = (\mathcal{Z}'\Omega^{-1}\mathcal{Z})^{-1}$ and $\Omega = \text{diag}[\omega_q^2, q = 1, \dots, Q]$.

(iii) Drawing from the conditional posterior of Σ

Using standard results for the Wishart distribution (Zellner, 1971, p. 227) the posterior conditional can be written as

$$p(\Sigma|\beta, u_1, \tilde{\mathbf{w}}, \mathcal{D}) \propto |\Sigma|^{\{N - (J+1) + N\}/2} \exp\left\{-\frac{1}{2}\text{tr}\Sigma^{-1}(\underline{\mathbf{S}} + \mathbf{S})\right\}, \quad (46)$$

where N is the total number of observations, $\mathbf{S} = \sum_{i,t} \hat{v}_{it}(\beta)\hat{v}_{it}(\beta)'$, and

$$\hat{v}_{it}(\beta) \equiv \begin{bmatrix} x_{1,it} - \mathbf{X}'_{it}\beta - u_{1,it} \\ \tilde{\mathbf{w}}_{it} - h(\mathbf{X}_{it}, \beta) \end{bmatrix}. \quad (47)$$

(iv) Drawing from the conditional posterior of Ω

The elements of the diagonal matrix Ω can be drawn easily as:

$$\frac{Q_q + (z_q - \mathcal{Z}\delta_q - \eta_q)'(z_q - \mathcal{Z}\delta_q - \eta_q)}{\omega_q^2} | \delta, \eta, \mathcal{D} \sim \chi^2(N + \underline{N}_q). \quad (48)$$

(v) Drawing from the conditional posterior of σ_u

This parameter can be drawn easily using the following:

$$\frac{Q_u + \mathbf{u}'_1\mathbf{u}_1}{\sigma_u^2} | \mathbf{u}, \mathcal{D} \sim \chi^2(N + \underline{N}_u). \quad (49)$$

(vi) Drawing from the conditional posterior of \mathbf{u}_1

We use the assumption that \mathbf{v}_1 and $\tilde{\mathbf{v}}$ are independent. To estimate inefficiency we examine the conditional distribution of \mathbf{u}_1 which is a truncated normal:

$$u_{1,it} | \boldsymbol{\beta}, \sigma_v, \sigma_u, \mathcal{D} \sim N_+ \left(\frac{x_{1,it} - \mathbf{X}'_{it} \boldsymbol{\beta}}{\sigma_u} \sigma_v^2, \frac{\sigma_v^2 \sigma_u^2}{\sigma_v^2 + \sigma_u^2} \right), \quad (50)$$

where σ_v^2 is the (1,1) element of $\boldsymbol{\Sigma}$. This conditional distribution is used explicitly in MCMC to generate draws for $u_{1,it}$. Given MCMC draws $\{u_{1,it}^{(s)}, s = 1, \dots, S\}$, efficiency draws are generated as $r_{1,it}^{(s)} = \exp(-u_{1,it}^{(s)})$, $s = 1, \dots, S$. The posterior mean of technical efficiency across all draws will be:

$$r_{1,it}^\eta = S^{-1} \sum_{s=1}^S \exp(-u_{1,it}^{(s)}). \quad (51)$$

(vii) Drawing from the conditional posterior of η

To estimate environmental efficiency for each bad output Z_q , we consider the conditional distribution of $\eta_{q,it}$, $q = 1, \dots, Q$ which is a truncated normal:

$$\eta_{q,it} | \boldsymbol{\delta}, \boldsymbol{\sigma}_\xi, \boldsymbol{\sigma}_\eta, \mathcal{D} \sim N_+ \left(\frac{z_{q,it} - \mathbf{Z}'_{it} \boldsymbol{\delta}_q}{\boldsymbol{\sigma}_{\xi,q}} \boldsymbol{\sigma}_{\eta,q}^2, \frac{\boldsymbol{\sigma}_{\xi,q}^2 \boldsymbol{\sigma}_{\eta,q}^2}{\boldsymbol{\sigma}_{\xi,q}^2 + \boldsymbol{\sigma}_{\eta,q}^2} \right). \quad (52)$$

This conditional distribution is used explicitly in MCMC when we draw from it to generate draws for $\eta_{q,it}$. Suppose the MCMC draws are $\{\eta_{q,it}^{(s)}, s = 1, \dots, S\}$. Then the posterior mean of environmental efficiency associated with z_q , $s = 1, \dots, S$ across all draws will be:

$$r_{q,it}^\eta = S^{-1} \sum_{s=1}^S \exp(-\eta_{q,it}^{(s)}), \quad (53)$$

which can be taken as a predictor of environmental efficiency for bad output q and observation corresponding to (i, t) . A kernel density of $r_{q,it}^\eta$ across observations for all $q = 1, \dots, Q$ then provides the sample distribution of our best predictor for environmental efficiency.

(viii) Drawing from the conditional posterior of $\tilde{\mathbf{w}}$

This step involves drawing from the following conditional posterior distributions:

(viii.1) Draw $\sigma_\lambda, \sigma_\mu, \boldsymbol{\sigma}_\varepsilon$ conditional on all other parameters below. Notice that $\boldsymbol{\sigma}_\varepsilon = [\sigma_{\varepsilon_j}^2, j = 2, \dots, J]$.

(viii.2) Draw from $\boldsymbol{\lambda} = [\lambda_i, i = 1, \dots, n] | \boldsymbol{\vartheta}, \sigma_\lambda, \sigma_\mu, \sigma_\omega, \boldsymbol{\mu}, \mathbf{a}, \mathcal{D}$, where $\mathbf{a} = [a_j, j = 2, \dots, J]$ and $\boldsymbol{\vartheta}$ denotes all parameters in steps (i) through (vii).

(viii.3) Draw from $\boldsymbol{\mu} = [\mu_t, t = 1, \dots, T] | \boldsymbol{\vartheta}, \sigma_\lambda, \sigma_\mu, \sigma_\omega, \boldsymbol{\lambda}, \mathbf{a}, \mathcal{D}$.

(viii.4) Draw $\tilde{\mathbf{w}} | \boldsymbol{\vartheta}, \sigma_\lambda, \sigma_\mu, \boldsymbol{\sigma}_\varepsilon, \boldsymbol{\lambda}, \boldsymbol{\mu}, \mathbf{a}, \mathcal{D}$.

The parameters in step (viii.1) can be drawn easily. Specifically, we have:

$$\frac{Q_\lambda + \boldsymbol{\lambda}' \boldsymbol{\lambda}}{\sigma_\lambda^2} | \boldsymbol{\lambda}, \mathcal{D} \sim \chi^2(n + \underline{N}_\lambda), \quad (54)$$

$$\frac{Q_\mu + \boldsymbol{\mu}' \boldsymbol{\mu}}{\sigma_\mu^2} | \boldsymbol{\mu}, \mathcal{D} \sim \chi^2(T + \underline{N}_\mu), \quad (55)$$

$$\frac{Q_{\tilde{w}} + \sum_{j,i,t} (\tilde{w}_{j,it} - a_j - \lambda_i - \mu_t)^2}{\sigma_\lambda^2} | \boldsymbol{\lambda}, \boldsymbol{\mu}, \tilde{\boldsymbol{w}}, \boldsymbol{a}, \mathcal{D} \sim \chi^2(N + \underline{N}_{\tilde{w}}). \quad (56)$$

The specification in (33) can be written as:

$$\tilde{\boldsymbol{w}}_{it} = \boldsymbol{a} + (\lambda_i + \mu_t) \mathbf{1}_{J-1} + \boldsymbol{\varepsilon}_{it}, \quad (57)$$

where $\boldsymbol{a} = [a_j, j = 2, \dots, J]$, $\mathbf{1}_{J-1}$ is the unit vector in \mathbb{R}^{J-1} and $\boldsymbol{\varepsilon}_{it} = [\varepsilon_{jit}, j = 2, \dots, J]$. In this form, parameters \boldsymbol{a} , $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ can be drawn jointly using standard results from linear models with random effects. Equation (57) can be written as:

$$\tilde{\boldsymbol{w}}_j = a_j \mathbf{1}_N + \left[\boldsymbol{D}_F : \boldsymbol{D}_T \right] \begin{bmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\mu} \end{bmatrix} + \boldsymbol{\varepsilon}_j, \quad j = 1, \dots, J, \quad (58)$$

where $\boldsymbol{D}_F, \boldsymbol{D}_T$ denote respectively firm and time dummy variables. Equivalently we can omit one firm-specific and one time-specific dummy (say the last) to obtain:

$$\tilde{\boldsymbol{w}}_j = a_j \mathbf{1}_N + \left[\boldsymbol{D}_F^* : \boldsymbol{D}_T^* \right] \begin{bmatrix} \boldsymbol{\lambda}^* \\ \boldsymbol{\mu}^* \end{bmatrix} + \boldsymbol{\varepsilon}_j, \quad j = 1, \dots, J, \quad (59)$$

where $\boldsymbol{\lambda}^*$ and $\boldsymbol{\mu}^*$ denote deviations of $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ from λ_n and μ_T respectively. Writing the equation above as:

$$\tilde{\boldsymbol{w}}_j = a_j \mathbf{1}_N + \boldsymbol{D}_o^* \boldsymbol{\psi} + \boldsymbol{\varepsilon}_j, \quad j = 1, \dots, J \quad (60)$$

and letting $\left[\mathbf{1}_N : \boldsymbol{D}_o^* \right] = \boldsymbol{D}^*$ we have

$$\tilde{\boldsymbol{W}} = \left[\boldsymbol{I}_{J-1} \otimes \mathbf{1}_N : \mathbf{1}_{J-1} \otimes \boldsymbol{D}_o^* \right] \begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{\psi} \end{bmatrix} + \boldsymbol{\varepsilon} \equiv \boldsymbol{\Delta} \boldsymbol{\varphi} + \boldsymbol{\varepsilon}, \quad (61)$$

where $\tilde{\boldsymbol{W}} = \left[\tilde{\boldsymbol{w}}_2 : \dots : \tilde{\boldsymbol{w}}_J \right]$. In this formulation the random effects can be drawn jointly from a normal distribution whose mean is $\hat{\boldsymbol{\varphi}} = (\boldsymbol{\Delta}' \boldsymbol{V}^{-1} \boldsymbol{\Delta})^{-1} \boldsymbol{\Delta}' \boldsymbol{V}^{-1} \tilde{\boldsymbol{W}}$ and the covariance is $(\boldsymbol{\Delta}' \boldsymbol{V}^{-1} \boldsymbol{\Delta})^{-1}$, where $\boldsymbol{V} = \text{cov}(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma}_\varepsilon \otimes (\boldsymbol{I}_{J-1} \otimes \boldsymbol{I}_N)$ and $\boldsymbol{\Sigma}_\varepsilon = \text{diag}[\sigma_{\varepsilon_j}^2, j = 2, \dots, J]$

In step (viii.4) we combine (57) with (28) and we write the result compactly as follows:

$$\begin{aligned} \tilde{\boldsymbol{w}}_{it} &= \boldsymbol{R}_{it} + \boldsymbol{\varepsilon}_{it} \\ \tilde{\boldsymbol{w}}_{it} &= h(\boldsymbol{X}_{it}, \boldsymbol{\beta}) + \tilde{\boldsymbol{v}}_{it} \end{aligned}$$

where $\boldsymbol{R}_{it} = \boldsymbol{a} + (\lambda_i + \mu_t) \mathbf{1}_{J-1}$ from (57). Denote $\tilde{\boldsymbol{\Sigma}} = \text{cov}(\tilde{\boldsymbol{v}}_{it})$, the lower $(J-1) \times (J-1)$ submatrix of $\boldsymbol{\Sigma}$. It is easy to show that the conditional posterior distribution of $\tilde{\boldsymbol{w}}_{it}$ is normal with mean $m_{it} = \frac{\mathbf{1}'_{J-1} \boldsymbol{\Sigma}_\varepsilon^{-1} \boldsymbol{R}_{it} + \mathbf{1}'_{J-1} \tilde{\boldsymbol{\Sigma}}^{-1} \boldsymbol{h}_{it}}{\mathbf{1}'_{J-1} (\boldsymbol{\Sigma}_\varepsilon^{-1} + \tilde{\boldsymbol{\Sigma}}^{-1}) \mathbf{1}_{J-1}}$ and variance $s_{it}^2 = \frac{1}{\mathbf{1}'_{J-1} (\boldsymbol{\Sigma}_\varepsilon^{-1} + \tilde{\boldsymbol{\Sigma}}^{-1}) \mathbf{1}_{J-1}}$, where $\boldsymbol{h}_{it} \equiv h(\boldsymbol{X}_{it}, \boldsymbol{\beta})$. Of course, drawing from this conditional posterior distribution is particularly easy. Finally, drawing from the conditional posterior distribution of $\boldsymbol{\sigma}_\varepsilon$ relies on the following:

$$\frac{Q_{\varepsilon_j} + \sum_{i,t} (\tilde{w}_{j,it} - R_{j,it})^2}{\sigma_{\varepsilon_j}^2} \sim \chi^2(N + \underline{N}_{\varepsilon_j}), \quad j = 2, \dots, J. \quad (62)$$