Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Journal of Econometrics 126 (2005) 445–468
JOURNAL OF Econometrics
www.elsevier.com/locate/econbase

Bayesian measurement of productivity and efficiency in the presence of undesirable outputs: crediting electric utilities for reducing air pollution

Scott E. Atkinson[a,*], Jeffrey H. Dorfman[b]

[a]*Department of Economics, University of Georgia, Athens, GA 30602, USA*
[b]*Department of Agricultural and Applied Economics, University of Georgia, Athens, GA 30602, USA*

Available online 26 June 2004

Abstract

Many studies have measured productivity change and efficiency when an undesirable output is a by-product. We flexibly treat the bad as a technology shifter of an input distance function and model a system of nonlinear equations subject to endogeneity. Theory dictates that we impose monotonicity on all inputs, outputs, and the bad. Since a Bayesian full-information likelihood approach can easily be misspecified, we utilize the Kim (J. Econometrics 107 (2002) 175) limited-information likelihood (LIL) derived by minimizing the entropy distance subject to the moment conditions from the Generalized Method of Moments (GMM) estimator. This represents an extension of the Bayesian Method of Moments approach of Zellner and Chen (Macroeconom. Dyn. 5 (2001) 673), Zellner and Tobias (Int. Econom. Rev. 42 (2001) 121), and Zellner (in: Bayesian Analysis in Econometrics and Statistics: The Zellner View and Papers, Edward Elgar, Cheltenham, 1997; J. Econometrics 83 (1998) 185) which uses entropy maximization but does not incorporate a specific likelihood. Using Bayes' Theorem we combine traditional priors with the LIL, which has a mode at the standard multiple-equation GMM estimator, yielding a limited-information posterior distribution. We generalize the approach of Kim (J. Econometrics 107 (2002) 175) by incorporating an unknown covariance matrix in a Gibbs sampling framework and applying the methodology to nonlinear equations. This allows us to estimate shadow prices, technical efficiency, and productivity change for a

*Corresponding author. Fax: +1-706-542-3376.
E-mail addresses: atknsn@terry.uga.edu (S.E. Atkinson), jdorfman@agecon.uga.edu (J.H. Dorfman).

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2004.05.009

panel of electric utilities, yielding results that differ substantially from those obtained using standard GMM.

JEL classification: C11; C32; D24

Keywords: Gibbs sampling; Distance functions; Productivity change; Technical efficiency rankings; Generalized method of moments

1. Introduction

A growing literature examines the measurement of productivity change (PC) and efficiency in industries which produce an undesirable output (a bad) as a by-product of their production processes. Historically, bads have been ignored in modelling production functions or cost functions for regulated industries. However, the emerging literature focuses on how to credit firms for reductions in bads that occur either at the expense of decreased production of desirable outputs or increased usage of inputs.

The non-stochastic approach is based on Pittman (1983), who extended the Caves et al. (1982) (hereafter, CCD) Malmquist indices of PC for multiple-input, multiple-output technologies. Their output-based distance measure of PC calculates the maximum proportional increase in desirable outputs (goods) for a given level of inputs. Their input-based distance measure of PC computes the maximum proportional decrease in inputs for a given level of goods.

Pittman (1983) realized that the CCD measures, while valid when all outputs are goods, were not directly applicable if some outputs were bads, such as pollution. By diverting inputs from reduction of bads to the production of goods, one could increase both goods and bads by a given proportion without any increase in input usage. Using the CCD definition of an output distance function, positive PC would have occurred, since more of both sets of outputs would have been produced with the same level of inputs.

Pittman (1983) proposed measuring the maximal radial expansion of goods and contraction of bads, holding inputs constant. However, this measure is no longer a distance function, since goods and bads are not scaled by the same radial expansion factor. Färe et al. (1989) used non-linear programming techniques to construct hyperbolic efficiency measures as proposed by Pittman.[1] Yaisawarng and Klein (1994) extended the Färe et al. techniques to measure the effect of controlling sulfur dioxide (SO_2) emissions on PC using a DEA model.

[1]In one formulation, inputs and bads are divided by the same scalar that multiplies the goods. That is, their approach treats the bad like an input. However, this imposes the restriction that the bad is a linearly homogeneous function of inputs, since both are scaled by the same factor. With the exception of Barbera and McConnell (1990), previous studies do not appear to recognize the functional relationship between emissions and inputs.

More recently, Chambers et al. (1996, 1998), and Färe et al. (1997) have used the directional distance function to calculate production relationships involving goods and bads. This approach provides a difference-based measure of PC rather than a ratio-based measure as with the Malmquist index based on the distance function.

The stochastic approach to measurement of productivity growth when a firm is producing goods and bads has most commonly been carried out using a cost function approach.[2] These studies can be separated into two basic approaches. The first has been to disaggregate a subset of inputs into abatement and non-abatement components in order to calculate their effect on costs. Barbera and McConnell (1990) estimated a translog cost function whose arguments were electricity output, the prices of capital, labor, energy, and materials, and the quantity of abatement capital. These authors compute the marginal reduction in PC due to abatement capital as the marginal effect of abatement capital on total costs. However, their approach does not consider the abatement components of the other inputs. They estimate that investment in pollution abatement capital in five manufacturing industries accounted for 10–30% of the decline in PC during the 1970s.[3]

The second approach has been to reformulate the bad as a good, namely, the percent reduction in the bad, thereby avoiding the problem raised by Pittman. Gollop and Roberts (1983) estimate that pollution control by environmentally constrained utilities reduced PC by 44% from 1973 to 1979. However, when computing the percent reduction in emissions, each choice of the base unconstrained emission rate creates a different nonlinear transformation of the original variable and hence a new model with different elasticities, returns to scale, and test statistics. Thus, it is important to implement a stochastic method for incorporating direct measures of bads themselves.[4]

Recently, productivity change with undesirable outputs has also been estimated using stochastic production frontiers by Fernandez et al. (2000), Koop (1998), and Reinhard and Thijssen (1998). However, each of these studies makes certain assumptions, such as separability of the technology in order to aggregate outputs and inputs, which may not be fully consistent with the empirical facts.

While some studies treat the bad as an input and others redefine the bad as a good, in this paper we are less restrictive by treating the bad as an ‘‘exogeneous’’ technology shifter and estimating an input distance system. We impose monotonicity on all inputs, outputs, and the bad for a distance system comprised of a distance equation and a set of nonlinear price equations, subject to endogeneity. However, since a Bayesian approach utilizing a full-information likelihood can easily be

[2]A number of papers estimating PC in the electric utility industry, such as Nelson (1984), Baltagi and Griffin (1988), and Callan (1991), ignored pollution as an output, even though inputs used to control pollution are included.

[3]A similar study by Conrad and Wastl (1995) disaggregated only materials into abatement and non-abatement components and finds reductions in PC as great as 15% due to abatement materials.

[4]A general problem with both of these stochastic approaches is their use of the Divisia index, where share-weighted averages of the proportional rates of growth of inputs are subtracted from a similarly constructed index of the outputs. This requires the restrictive assumption of constant returns to scale, which is inconsistent with the unrestricted translog cost function employed in each study.

misspecified, we wish to employ a limited-information likelihood (LIL). Kim (2002) constructs a LIL using the moment conditions of GMM while minimizing the entropy distance. That is, his approach attempts to minimize the chance of misspecification by considering a set of LIL functions that satisfy the GMM moment conditions on the parameters and choosing from this set the LIL function that has its mode at the standard GMM estimator and is the closest to the true likelihood in the Kullback–Leibler information criterion distance or entropy distance. The LIL function can be directly combined with traditional priors to obtain a limited-information posterior (LIP) for all parameters of interest using Bayes' Theorem. This GMM limited-information approach allows us to easily deal with endogeneity, which we suspect is an important feature of our data.

An alternative approach has been pursued by Zellner and Tobias (2001), Zellner and Chen (2001), and Zellner (1997, 1998). Without incorporating a specific likelihood, they develop a Bayesian Method of Moments (BMOM) estimator that maximizes entropy subject to a set of moment conditions to obtain a "post-data" density. Zellner (1997) and Zellner and Tobias (2001) state that their post-data density can be combined with priors to obtain a LIP density. Further, Kim (2002) shows that his LIL based on the GMM moment conditions is an extension of their post-data density to the case of GMM.

Therefore, we apply the Kim limited-information GMM approach to a system of nonlinear equations, which is also the focus of the BMOM models of Zellner and Chen (2001) and Zellner (1998). In so doing, we generalize the approach of Kim (2002) to the case of an unknown covariance matrix using Gibbs sampling to analyze the LIP and derive empirical results that demonstrate the usefulness of the new methodology. Based on an application of our limited-information estimator to a panel of electric utilities, the Bayesian estimator of the shadow price of the bad is much more accurate than its non-Bayesian counterpart. Further, productivity change and technical change are substantially higher when the bad is included.

The remainder of this paper is organized as follows. In Section 2, we present our distance function specification that includes bads. In Section 3, we present our parametric specification. Section 4 presents background and derivation of the estimator we use in our application. Section 5 discusses data, further details of our estimation procedure, and the results of our empirical application. Conclusions follow in Section 6.

2. Distance function specifications

One can formulate a number of stochastic distance function specifications incorporating bads. The caveat is that one cannot treat bads symmetrically with either the inputs or outputs that are being scaled radially. As shown by Pittman (1983), a misspecification results if we scale bads proportionally with goods in an output distance function. If we treat bads symmetrically with inputs, an input distance function would scale both proportionally, imposing the restrictive assumption that the pollution generating function is linearly homogeneous in input quantities.

Our preferred approach is to estimate an input distance function where the bad is assumed to be quasi-fixed as an exogenous shifter of the isoquant (but not treated symmetrically with goods) in the short run. In the long run we allow the bad to vary in order to compute its shadow price.

2.1. Distance functions ignoring bads

Let $\mathbf{x}$ be a vector of inputs $\mathbf{x} = (x_1, \ldots, x_N) \in R_+^N$ and let $\mathbf{y}$ be a vector of goods denoted by $\mathbf{y} = (y_1, \ldots, y_M) \in R_+^M$. Disregarding bads, one can write the production technology, $S(\mathbf{x}, \mathbf{y}, t)$, as

$$S(\mathbf{x}, \mathbf{y}, t) = \{(\mathbf{x}, \mathbf{y}) : \ \mathbf{x} \text{ can produce } \mathbf{y} \text{ at time } t\}, \tag{2.1}$$

where $t = 1, \ldots, T$ is time. The output distance function is defined as

$$D_o(\mathbf{x}, \mathbf{y}, t) = \inf_{\theta}\{\theta : \ (\mathbf{x}, \mathbf{y}/\theta) \in S(\mathbf{x}, \mathbf{y}, t)\} \tag{2.2}$$

and the input distance function is defined as

$$D_i(\mathbf{y}, \mathbf{x}, t) = \sup_{\lambda}\{\lambda : \ (\mathbf{x}/\lambda, \mathbf{y}) \in S(\mathbf{x}, \mathbf{y}, t)\}, \tag{2.3}$$

where we follow the notational convention that each distance function is linearly homogeneous with respect to the last-named input or output vector (or vectors if in square brackets). Clearly, both (2.2) and (2.3) are misspecifications if the bad is correlated with inputs or goods.

2.2. Input distance functions: bads as technology shifters

We now include a vector of bads $\mathbf{b}$, $(b_1, \ldots, b_Z) \in R_+^Z$, which is produced jointly with $\mathbf{y}$. The production of bads is a function of the inputs, the goods, and the state of technology:

$$\mathbf{b} = \mathbf{b}(\mathbf{x}, \mathbf{y}, t). \tag{2.4}$$

Symmetric treatment of bads and goods using an input distance function can be specified as

$$D_i([\mathbf{y}, \mathbf{b}], \mathbf{x}, t) = \sup_{\lambda}\{\lambda : \ ([\mathbf{y}, \mathbf{b}], \mathbf{x}/\lambda) \in S(\mathbf{x}, \mathbf{b}, \mathbf{y}, t)\}. \tag{2.5}$$

Here the goods and bads are held constant and inputs are proportionally scaled downward to their minimum required level.

Since the input distance function in (2.5) is dual to the cost function, we can write

$$C([\mathbf{y}, \mathbf{b}], \mathbf{p}, t) = \min_{\mathbf{x}}\{\mathbf{px} : \ D_i([\mathbf{y}, \mathbf{b}], \mathbf{x}, t) \geqslant 1\}, \tag{2.6}$$

where $\mathbf{p} = (\mathbf{p}_1, \ldots, \mathbf{p}_N) \in R_+^N$ is a vector of input prices and $C([\mathbf{y}, \mathbf{b}], \mathbf{p}, t)$ is a cost function. This equation implies that unless inputs are used in their cost-minimizing proportions, the input distance measure will be greater than one. Formulating the associated Lagrangian and taking the first-order conditions, Färe and Primont (1995) show that the shadow value for each input is given by

$$\mathbf{p} = C([\mathbf{y},\mathbf{b}],\mathbf{p},t)\nabla_x D_i([\mathbf{y},\mathbf{b}],\mathbf{x},t), \tag{2.7}$$

where $C([\mathbf{y},\mathbf{b}],\mathbf{p},t)$ is the value of the Lagrange multiplier.

Equivalently, the bads can be treated as exogenous shifters of the technology set, similar to a time trend or state of technology variable. The notion here is that conditional on the level of the bad, efficiency measures over the desirable outputs and inputs are well-defined and behave as expected. Yet, ignoring the bads would lead to biased results since firms would not receive credit for input use that is directed at reduction of the bad. Treating the level of the bad as a shifter of the technology set allows firms to be credited (penalized) for reducing (increasing) the level of bad that they produce.

To emphasize this point, Eq. (2.5) can now be written as

$$D_i(\mathbf{y},\mathbf{x},t|\mathbf{b}) = \sup_{\lambda}\{\lambda:\ (\mathbf{x}/\lambda,\mathbf{y}|\mathbf{b}) \in S(\mathbf{x},\mathbf{y},t|\mathbf{b})\}. \tag{2.8}$$

A useful simplification for thinking about this approach is an isoquant for one good and two inputs. Ignoring the bad, we can draw isoquants in a standard two-dimensional space and measure efficiency as the radial distance from the production point to the origin relative to the distance on the same radial from the isoquant to the origin. For a firm with a lower level of the bad, the isoquant shifts outward reflecting that a higher combination of inputs is required to produce some level of the desirable output, recognizing that some input use is directed at reducing the production of the bad.

The appropriate monotonicity condition for the bad in the context of the input distance function can be derived as follows. The input distance function is monotonically nondecreasing in inputs ($\partial D_i/\partial x_n \geqslant 0$) and monotonically nonincreasing in outputs ($\partial D_i/\partial y_m \leqslant 0$). Assuming a single bad, we compute the partial total differential of Eq. (2.8) evaluated on the frontier at a fixed time [implying $D_i(\mathbf{y},\mathbf{x},t|\mathbf{b}) = 1$ and $\mathrm{d}t = 0$] to obtain

$$\mathrm{d}D_i = \sum \frac{\partial D_i}{\partial y_m}\,\mathrm{d}y_m + \sum \frac{\partial D_i}{\partial x_n}\,\mathrm{d}x_n + \frac{\partial D_i}{\partial b}\,\mathrm{d}b = 0. \tag{2.9}$$

In order to keep the firm on the input distance frontier, we set $\mathrm{d}y_m = 0$, $\forall m$ and obtain

$$\frac{\partial D_i}{\partial b} = -\sum \frac{\partial D_i}{\partial x_n}\frac{\mathrm{d}x_n}{\mathrm{d}b}. \tag{2.10}$$

To mirror the reality of ever-shrinking SO_2 emission allowances, we assume that the bad can only be decreased. Following Pittman (1983), with constant desirable output and technology, bads can only be reduced through increased usage of at least some inputs (such as using more labor, more capital, or switching to low-sulphur coal).[5]

[5]An exception to this would be the use of less high-sulphur coal, which would also reduce pollution. However, this would require the increased use of other inputs to maintain constant output.

This implies that $\mathrm{d}x_n/\mathrm{d}b \leqslant 0$ for most inputs, which combined with the nonnegativity property for inputs, $\partial D_i/\partial x_n \geqslant 0$, should yield $\sum(\partial D_i/\partial x_n)(\mathrm{d}x_n/\mathrm{d}b) \leqslant 0$ and $(\partial D_i/\partial b) \geqslant 0$.

3. Parametric specification

As a flexible approximation to the true distance function in (2.3), we adopt the translog functional form. Thus, the empirical model for firm $f = 1, \ldots, F$ in period $t = 1, \ldots, T$ has the form

$$\begin{aligned}
0 = {} & \gamma_0 + \sum_m \gamma_m \ln y_{mft} + \sum_z \gamma_z \ln b_{zft} \\
& + \sum_n \gamma_n \ln x_{nft} + \gamma_{t1} t + (1/2)\gamma_{t2} t^2 \\
& + (1/2) \sum_m \sum_{m'} \gamma_{mm'} \ln y_{mft} \ln y_{m'ft} + (1/2) \sum_z \sum_{z'} \gamma_{zz'} \ln b_{zft} \ln b_{z'ft} \\
& + (1/2) \sum_n \sum_{n'} \gamma_{nn'} \ln x_{nft} \ln x_{n'ft} + \sum_m \sum_n \gamma_{mn} \ln y_{mft} \ln x_{nft} \\
& + \sum_z \sum_n \gamma_{zn} \ln b_{zft} \ln x_{nft} + \sum_z \sum_m \gamma_{zm} \ln b_{zft} \ln y_{mft} \\
& + \sum_m \gamma_{mt} \ln y_{mft} t + \sum_z \gamma_{zt} \ln b_{zft} t \\
& + \sum_n \gamma_{nt} \ln x_{nft} t + \ln h(\varepsilon_{ft}),
\end{aligned} \tag{3.1}$$

where

$$h(\varepsilon_{ft}) = \exp(v_{ft} - u_{ft}), \tag{3.2}$$

so that $\ln h(\varepsilon_{ft})$ is an additive error with a one-sided component, u_{ft}, and a standard noise component, v_{ft}, with zero mean.[6] We model the unobserved heterogeneity, u_{ft}, using a fixed-effects specification. To identify the u_{ft} for each f and t, we require that additional restrictions be imposed on the pattern of technical efficiency over time. Using the model for time-varying inefficiency proposed by Cornwell et al. (1990), we examine specifications of the form

$$u_{ft} = \sum_{q=0}^{Q} \beta_{fq} d_f t^q, \quad f = 1, \ldots, F, \tag{3.3}$$

where t is a trend and d_f is a dummy variable equal to one for firm f and zero for the other firms. The estimated version of (3.1) is obtained by substituting (3.3) into (3.2), which in turn is substituted into (3.1), so that the β_{fq} are firm-specific parameters to

[6]Since the inclusion of v_{ft} makes the frontier distance function stochastic, it is possible for $h(\varepsilon_{ft})$ to be greater than 1.

be estimated. If identification becomes difficult with large F, alternative fixed-effects models can be employed.[7] However, the fixed-effects approach avoids the distributional and exogeneity assumptions that would otherwise be required in a random effects setup.[8]

We also use (2.7) to derive price equations for each input and append i.i.d. error terms to these equations. These price equations together with the distance equation (3.1) comprise our estimated distance system.

Prior to estimation, several sets of parametric restrictions are imposed on (3.1). Symmetry requires that

$$\begin{aligned} &\gamma_{mm'} = \gamma_{m'm}, \quad \forall m, m', \ m \neq m', \\ &\gamma_{zz'} = \gamma_{z'z}, \quad \forall z, z', \ z \neq z', \\ &\gamma_{nn'} = \gamma_{n'n}, \quad \forall n, n', \ n \neq n'. \end{aligned} \tag{3.4}$$

In addition, linear homogeneity in input quantities implies

$$\begin{aligned} &\sum_n \gamma_n = 1, \\ &\sum_n \gamma_{nn'} = \sum_{n'} \gamma_{nn'} = \sum_n \sum_{n'} \gamma_{nn'} = 0, \\ &\sum_n \gamma_{mn} = 0, \quad \forall m, \\ &\sum_n \gamma_{zn} = 0, \quad \forall z \end{aligned}$$

and

$$\sum_n \gamma_{nt} = 0. \tag{3.5}$$

Finally, identification requires that β_{fq}, $\forall q$, must be constrained for one firm in (3.3).

Given the cross-equation restrictions implied by the symmetry and linear homogeneity restrictions, we achieve gains in efficiency by estimating the distance equation in (3.1), subject to (3.2) and (3.3), together with (2.7) for each input, for a total of four equations. We then compute levels of technical efficiency, efficiency change (EC), technical change (TC), and PC. Since we do not impose one-sidedness (non-negativity) on the u_{ft} in estimation, we need to do so after estimation, by adding

[7] Even with F of 1,000 or more many econometrics packages can compute a dummy variable model with a unique parameter for each firm. When larger models become intractable, one could estimate time-demeaned or first-differenced models. However, the time-demeaned model differs from the dummy variable model, since our system is nonlinear in the parameters; the models are equivalent only with linear-in-parameters systems.

[8] If these assumptions hold, one could estimate a random effects model. If a GLS approach is taken, one would have to compute the variances of the unobserved heterogeneity and the idiosyncratic error. If a maximum likelihood approach is taken, one must specify the distribution of both errors, choosing some type of a one-sided distribution for the former. Koop (2001) parameterizes the mean of an exponential technical inefficiency distribution using a vector of dummy variables that proxy the state of the economy in the countries whose manufacturing sectors he is examining.

and subtracting from the fitted model $\hat{u}_t = \min_f(\hat{u}_{ft})$, which defines the frontier intercept. With $\ln \hat{D}_i(\mathbf{y}, \mathbf{x}, t)$ representing the estimated translog portion of (3.1) (i.e., those terms other than $h(\varepsilon_{ft})$), adding and subtracting $\hat{u}_t$ yields

$$0 = \ln \hat{D}_i(\mathbf{y}, \mathbf{x}, t) + \hat{v}_{ft} - \hat{u}_{ft} + \hat{u}_t - \hat{u}_t = \ln \hat{D}_i^*(\mathbf{y}, \mathbf{x}, t) + \hat{v}_{ft} - \hat{u}_{ft}^*, \tag{3.6}$$

where $\ln \hat{D}_i^*(\mathbf{y}, \mathbf{x}, t) = \ln \hat{D}_i(\mathbf{y}, \mathbf{x}, t) - \hat{u}_t$ is the estimated frontier distance function in period t and $\hat{u}_{ft}^* = \hat{u}_{ft} - \hat{u}_t \geqslant 0$.

Using (3.3), we estimate firm f's level of technical efficiency in period t, TE_{ft}, as

$$\mathrm{TE}_{ft} = \exp(-\hat{u}_{ft}^*), \tag{3.7}$$

where our normalization of $\hat{u}_{ft}^*$ guarantees that $0 < \mathrm{TE}_{ft} \leqslant 1$. Given an estimate of TE_{ft} obtained from (3.7), we then calculate EC_{ft} as the change in technical efficiency:

$$\mathrm{EC}_{ft} = \Delta\mathrm{TE}_{ft} = \mathrm{TE}_{ft} - \mathrm{TE}_{f,t-1}. \tag{3.8}$$

We measure TC as the difference between the estimated frontier distance function in periods t and $t-1$ holding output and input quantities constant:

$$\begin{aligned} \mathrm{TC}_{ft} &= \ln \hat{D}_i^*(\mathbf{y}, \mathbf{x}, t) - \ln \hat{D}_i^*(\mathbf{y}, \mathbf{x}, t-1) \\ &= \sum_m \hat{\gamma}_{mt} \ln y_{mft} + \sum_n \hat{\gamma}_{nt} \ln x_{nft} + \sum_z \hat{\gamma}_{zt} \ln b_{zft} \\ &\quad + \hat{\gamma}_{t1} + (1/2)\hat{\gamma}_{2t}[(t+1)^2 - t^2] - (\hat{u}_t - \hat{u}_{t-1}). \end{aligned} \tag{3.9}$$

Thus, the change in the frontier intercept, $\hat{u}_t$, affects TC as well as EC. Finally, given EC and TC, we follow Atkinson and Cornwell (1998) by constructing estimates of PC = TC + EC.

To estimate the shadow price for SO_2 emissions, we apply the envelope theorem to the Lagrangian corresponding to (2.6). We then use the fact that the Lagrange multiplier equals $C([\mathbf{y}, \mathbf{b}], \mathbf{p}, t)$, as shown in Färe and Primont (1995), to obtain

$$\frac{\partial C([\mathbf{y}, \mathbf{b}], \mathbf{p}, t)}{\partial \mathbf{b}} = -C([\mathbf{y}, \mathbf{b}], \mathbf{p}, t)\nabla_b D_i([\mathbf{y}, \mathbf{b}], \mathbf{x}([\mathbf{y}, \mathbf{b}], \mathbf{p}, t), t). \tag{3.10}$$

This expression can be interpreted as the shadow price of SO_2 emissions.

4. Background and motivation for the estimator

In this section we provide some background and motivation for the Limited Information Bayesian System Estimator (LIBSE), which we will use in the estimation of our distance system. This multiple equation system is nonlinear in the parameters and will require correction for endogeneity and the imposition of monotonicity restrictions.

One could employ GMM or full-information maximum likelihood (FIML) to estimate a nonlinear, multiple equation system. In the following application, FIML would involve dealing with four equations (three of which are nonlinear) and specifying additional structural equations for all endogenous variables in all four equations. In a Bayesian, full-information approach, identifying the parameters would require either zero restrictions or other forms of prior information (cf. Zellner, 1971, Chapter 9). The common multimodality of such posterior distributions led to the creation of MELO estimators by Zellner (1978), which have better-behaved posterior moments. Specification is complicated further when models must incorporate autocorrelation and heteroskedasticity. In traditional Bayesian analysis, these features must be fully specified in the likelihood function. However, in a world of limited information about the specification of the structural equations and the distribution of the dependent variables, misspecification is likely, in which case the desirable properties of full information analysis are lost.

The alternative taken here is to follow Kim (2002), which extends the BMOM approach of Zellner and Tobias (2001), Zellner and Chen (2001), and Zellner (1997, 1998).[9] Kim (2002) shows that maximizing entropy subject to a restriction on the GMM criterion function yields an optimal LIL of the exponential family and goes on to develop the apparatus necessary to embed GMM estimators in a Bayesian inference framework. We generalize Kim's approach by treating both the covariance of the errors and the unknown parameters of our distance system as random variables and then constructing a joint LIL for both. To accomplish this, we borrow from Zellner and Tobias (2001) by imposing moment conditions on both sets of parameters before solving for the joint LIL. In the process we extend the Zellner and Chen (2001) BMOM estimation of multiple equation models to nonlinear systems within the Bayesian LIL framework. We combine this LIL with a prior density to obtain joint and conditional LIP distributions for the distance system parameters and the covariance matrix. The conditional LIP will allow us to use Gibbs sampling in our application to numerically investigate posterior measures of the model parameters and functions of those parameters. Thus, we present a method for deriving a LIBSE which requires only specifying a set of GMM overidentifying restrictions and any prior densities that a researcher wishes to use.

Following the development in Kim (2002), a standard first-moment condition for an estimator γ is

$$\mathrm{E}[h(\gamma)|\Omega, D] = 0, \tag{4.1}$$

where a vector-valued function $h: \mathbb{R}^k \to \mathbb{R}^k$ operates on γ, a vector of regression model parameters, Ω is the covariance matrix of the regression model's stochastic error terms, D is the data (including instruments), and k is the number of exogenous variables in the system, equal to the dimension of $h(\gamma)$. Later, we will need some moment conditions on Ω in order to treat its elements as parameters to be estimated rather than known. The second moment condition is

$$\mathrm{E}[h(\gamma)h(\gamma)'|\Omega, D] = S, \tag{4.2}$$

where S is the covariance matrix of h.

[9] Zellner and Highfield (1988), Green and Strawderman (1996), and LaFrance (1999) are other applications of BMOM.

To embed these two sets of moment conditions in an inference framework, we follow Kim (2002, Eq. (3.1)) and define a set of admissible LIL functions by

$$\mathscr{F} = \{\ell(\gamma|\Omega, D):\ \mathrm{E}[h(\gamma)'S^{-1}h(\gamma)|\Omega, D] = k\}. \tag{4.3}$$

Note that the expected value of the criterion function in (4.3) is determined by using the second moment condition.

To choose among the set of admissible LIL functions, one must decide the selection criterion to employ. In the spirit of GMM estimators, it makes sense to choose the least informative (most diffuse) LIL function since that allows the researcher to proceed with the minimum amount of additional structure imposed on the estimation procedure. The least informative LIL function can be found by solving the following optimization problem

$$\underset{\ell \in \mathscr{F}}{\operatorname{argmax}} - \int \ell(\gamma|\Omega, D) \ln[\ell(\gamma|\Omega, D)]\, \mathrm{d}\gamma. \tag{4.4}$$

The optimization problem yields the probability measure that is closest to the true measure in terms of entropy distance, or equivalently, in terms of the Kullback-Liebler information criterion distance (Kim, 2002). This is analogous to the Zellner and Tobias (2001) maximization of entropy subject to a set of moment conditions where restricting the choice of densities to those in $\mathscr{F}$ is equivalent to their moment restrictions. The solution to the optimization problem in Eq. (4.4) is given by

$$\hat{\ell}(\gamma|\Omega, D) = c_o \exp[-c_1 h(\gamma)'S^{-1}h(\gamma)], \tag{4.5}$$

where c_o and c_1 are constants that control the scale and shape of the density and ensure that it is proper. The form of the optimal LIL function is a member of the exponential family of distributions.[10] The density in (4.5) can be used in place of the standard likelihood function since it is Borel measurable and proper, thereby allowing the use of Bayes' Theorem in later deriving a LIP distribution for the parameters of interest. The LIL function is also robust to misspecification, since it is based on minimal assumptions.

As in any Bayesian inference problem, in the final step, a prior density $p(\gamma)$ is combined with our LIL function using Bayes' Theorem to yield a LIP distribution

$$f(\gamma|\Omega, D) = p(\gamma)\hat{\ell}(\gamma|\Omega, D)c^{-1}, \tag{4.6}$$

where c is the normalizing constant.

The above development clearly demonstrates the validity of placing GMM-type estimators into an inference framework (Bayesian or sampling-theoretic). However, the equations we have presented (and the development in Kim (2002) which we have paralleled) treat Ω as fixed. This makes the LIL function in (4.5) similar to a concentrated or conditional likelihood. Zellner (1998) and Zellner and Tobias (2001)

[10] The non-Bayesian might note that if one used the LIL function in (4.5) as a likelihood function and solved for the value of γ that maximizes it (or its natural log), the resulting estimator is the standard GMM estimator of γ, demonstrating (as in Kim, 2002, Eq. (3.8)) that the GMM estimator can be viewed as a limited-information maximum-likelihood estimator, where the limitation on the information set concerns the degree of specificity in the likelihood function.

present moment conditions on single-equation variances that result in a variety of distributional forms for the LIP that are generalized forms of (4.6) above. In the following empirical application, sampling variation in the covariance matrix is important, since covariance matrix parameters are highly correlated with some of the regression parameters and estimates of the conditional and marginal posterior distributions of the regression parameters vary by (economically and statistically) significant amounts.

Thus, a joint LIP distribution for the covariance matrix and the regression parameters is desirable. We first obtain a joint LIL function by generalizing (4.3) to remove the conditioning on the covariance matrix employed by Kim (2002) in Eq. (3.1). Analogous to our (4.3), we define the new set of admissible LIL functions as

$$\mathscr{F} = \{\ell(\gamma, \Omega|D):\ \mathrm{E}[h(\gamma,\Omega)'S^{-1}h(\gamma,\Omega)|D] = k,\ \mathrm{E}[\mathrm{tr}(\Xi\Omega^{-1})] = \kappa,\ \mathrm{E}[\ln|\Omega|] = \tau\}, \tag{4.7}$$

where Ξ is the sum of squared residuals matrix, and κ and τ are scalar constants. Compared to (4.3), the two new moment restrictions on Ω produce a marginal LIL function that is an inverted Wishart with respect to Ω (Zellner and Tobias, 2001).

In conjunction with the choice of prior for Ω, appropriate choices of κ and τ, namely,

$$\kappa = \int \mathrm{tr}(\Xi\Omega^{-1})\exp\left[\frac{-(n-k)}{2}\ln|\Omega| - (1/2)\,\mathrm{tr}(\Xi\Omega^{-1})\right]\mathrm{d}\Omega \tag{4.8}$$

and

$$\tau = \int \ln|\Omega|\exp\left[\frac{-(n-k)}{2}\ln|\Omega| - (1/2)\,\mathrm{tr}(\Xi\Omega^{-1})\right]\mathrm{d}\Omega \tag{4.9}$$

yield an inverted Wishart in the posterior distribution with the standard parameters Ξ and $(n-k+m+1)/2$. Therefore, we set κ and τ to the above specified values.[11]

By solving a maximization problem analogous to (4.4) we obtain a joint LIL function analogous to (4.5):

$$\hat{\ell}(\gamma, \Omega|D) = c_o|\Omega|^{-c_1}\exp[-c_2 h(\gamma,\Omega)'S^{-1}h(\gamma,\Omega) - c_3\,\mathrm{tr}(\Xi\Omega^{-1})], \tag{4.10}$$

which is the product of a distribution from the exponential family for γ and an inverted Wishart with respect to Ω (Zellner and Tobias, 2001). The LIL function in (4.10) then can be combined with a prior density through Bayes' Theorem to derive a LIP analogous to (4.6).

[11]Alternatively, to fully maximize entropy one could estimate the parameters κ and τ of the inverted Wishart that maximize the entropy subject to the moment conditions in (4.7). One would be forced to perform a numerical search process involving the solution of (4.8) and (4.9). See Zellner and Highfield (1988). Also note that once Gibbs sampling is begun, Ω and Ξ would change throughout the numerical sampling process. This would require that these two equations be numerically resolved for each set of draws, which would be very time consuming. Given that interest in the parameters of the inverted Wishart is likely to be low, the tradeoff inherent in fixing these parameters a priori seems worthwhile.

5. Empirical application

We apply our methodology to a panel of U.S. electric utilities observed at 5-year intervals from 1980 to 1995. The good output is the quantity of electric power generated, the bad is SO_2 emissions (which locally has a direct negative effect on health and welfare and regionally can lead to acid rain), and the inputs are capital, labor, and energy. This application is particularly relevant since allowable SO_2 emissions from electric utilities have been reduced dramatically over the last decade. Title IV of the 1990 Clean Air Act Amendments reduced emissions of SO_2 from U.S. coal-burning electric utilities from about 19 million tons in 1980 to 8.95 million tons by the year 2000. The increased reduction of SO_2 emissions over time has likely had an important impact on the levels of technical efficiency and productivity growth for these utilities. Proper crediting for reduction of this bad is essential to obtain unbiased estimates of efficiency levels and productivity growth. It also can provide insights into what the tradeoff has been between emissions and output.

5.1. Data

Our data set is an updated and refined version of the panel of utilities originally analyzed by Nelson (1984).[12] Subsets of that data were used by Baltagi and Griffin (1988) and Callan (1991). The sample used here is comprised of 43 privately owned U.S. electric utilities for the years 1980, 1985, 1990, and 1995.[13] A list of the utilities and the firm number by which they are referenced henceforth in our tables is provided in Table 1. 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 fossil fuel-based steam electric generation.

Variable definitions for inputs quantities and prices as well as output quantities are generally consistent with those in Nelson (1984). The inputs are quantities of fuel (x_E), labor (x_L), and capital (x_K), measured as ratios of input expenditure to price. Electrical output (y) is defined as the sum of residential and industrial-commercial output in 10 millions of kilowatt hour sales and SO_2 emissions (b) are measured in tons. Details are available from the authors. The output observations compiled by Daniel McFadden and Thomas Cowing were updated using the *Statistics of Privately Owned Electric Utilities in the U.S.* Table 2 reports the average quantities of inputs and output. Over the 1980–1995 time period, x_K declined somewhat. More dramatic was the greater than 20% reduction in x_L and b. Finally, x_E and y increased moderately.

[12]We are grateful to Professor Nelson for making his data available to us.

[13]The primary sources for Nelson's sample are the U.S. Federal Power Commission's *Statistics of Privately Owned Electric Utilities in the U.S.*, *Steam Electric Plant Construction Cost and Annual Production Expenses*, and *Performance Profiles—Private Electric Utilities in the United States*: 1963–70. Additional data were taken from *Moody's Public Utility Manual.*

Table 1
Utilities in the sample

Firm number	Utility
1	Alabama PC
2	Arizona PSC
3	Arkansas PLC
4	Pacific GEC
5	SanDiego GEC
6	PSC Colorado
7	UIC Connecticut
8	Delmarva PLC
9	Potomac EPC
10	Tampa EC
11	Georgia PC
12	C Illinois PSC
13	PSC Indiana
14	PC Iowa
15	Kansas GEC
16	Kentucky UC
17	Louisville GEC
18	C Louisiana EC
19	C Maine PC
20	Baltimore GEC
21	Boston EC
22	Detroit EC
23	Mississippi PLC
24	Kansas City PLC
25	PSC New Hampshire
26	Atlantic City EC
27	PSEGC New Jersey
28	PSC New Mexico
29	Central Hudson GEC
30	GEC New York
31	Rochester GEC
32	Carolina PLC
33	Duke PC
34	Cleveland EIC
35	Ohio EC
36	Oklahoma GEC
37	DLC Pennsylvania
38	Philadelphia PC
39	West Penn PC
40	S Carolina EGC
41	Virginia EPC
42	Appalachian PC
43	Wisconsin EPC

Table 2
Average quantities of inputs, output, and emissions by year (units in parentheses)

Year	x_K (\$[a])	x_L (workers[a])	x_E (Million MMBtu)	b (tons[a])	y (Mwh[a])
1980	262.2834	0.1373	148.3840	13.2758	1407.6013
1985	205.9072	0.1516	155.8825	12.0573	1441.2334
1990	223.8414	0.1469	160.7528	12.9737	1515.3163
1995	253.7025	0.1094	159.6330	9.8837	1518.8712

[a]In units of 10,000.

Data on SO_2 emissions are published on the EPA Acid Rain Website.[14] The primary data are for Clean Air Act Amendment Phase I and Phase II units and were aggregated to the utility level. Whenever units were owned by more than one utility, emissions were allocated by ownership share. Emissions of SO_2 are measured in tons. Data on emissions are available for 1980, 1985, and 1990 as historical EPA estimates, while the 1995 data are actual (measured) emissions from EPA's Continuous Emission Monitoring System. Thus, our panel is comprised of 43 firms for the years 1980, 1985, 1990, and 1995, for a total of 172 observations.

5.2. Bayesian estimation procedure

In total, we have a distance system comprised of four equations: the distance function and three first-order conditions, one for each of the three inputs. To implement the estimation in a LIBSE framework, we find the LIL function using the maxent approach incorporating moment conditions on the distance system parameters and the error covariances, using Eqs. (4.7) and (4.10). The priors employed are informative, diffuse but proper in form, with the addition of monotonicity restrictions on derivatives of the distance function. The LIL function is combined with the informative priors to derive a LIP analogous to (4.6). Our extensions of Kim (2002) and Zellner and Tobias (2001) allow the use of instruments to address the endogeneity inherent in estimation of a distance function and the incorporation of informative priors on the parameters while still yielding exact finite sample posterior moments for the parameters of interest (Zellner, 1998).

As indicated above, our estimated distance system consists of four equations—(3.1), subject to (3.2) and (3.3), and (2.7) for each input. We impose (3.4) and (3.5) in (3.1) and (2.7) and append an additive iid error term, w_{ft}^k, $k = 1, \ldots, 3$, to each equation in (2.7). With $Q = 2$, we use the standard GMM estimator to test a number of null hypotheses regarding (3.3). We employ a quasi-likelihood ratio test statistic that equals the sample size times the difference between the restricted and unrestricted criterion functions, which is asymptotically distributed as chi-square. At the 0.01 level we fail to reject the null hypothesis that $\beta_{f0} = 0$, $\forall f$ and

[14]http://www.epa.gov/acidrain/scorcard/es1995.html

subsequently drop the corresponding firm dummies. This result is not surprising, since we have only four years of data for each firm. Similarly, we fail to reject the null hypothesis that $\beta_{f2} = 0, \forall f$ and therefore Q is set equal to 1 in (3.3), for the final estimates. We also impose $\beta_{11} = 0$ to achieve identification.

5.2.1. Specification of the prior and derivation of the limited information posterior

In the specification of the prior, we differ from Zellner (1998) and Zellner and Tobias (2001), by going beyond the proper maxent prior to a more informative one. The full prior distribution will be a product of independent priors on the structural parameters of the distance function, the prior on the covariance matrix of the vector of errors, and a set of indicator functions that restrict prior support to the region where the monotonicity restrictions are satisfied. Denote the prior to be used in the analysis as

$$p(\gamma, \Omega) = p(\gamma)p(\Omega)\mathrm{I}(\gamma, \mathscr{R}). \tag{5.1}$$

The structural parameters, γ, of the distance function are each given a normal prior distribution with zero mean and variance of 100. This is a diffuse enough prior to have virtually no effect on the determination of the posterior means, but does ensure that the prior is proper in any dimensions that are not restricted to a finite subspace by the indicator function part of the prior. It also makes the LIP density straightforward to work with when we begin Gibbs sampling.[15] The prior for Ω (the matrix of variances and covariances of the four errors appended to the equations to be estimated) is a standard Jeffreys prior,

$$p(\Omega) \propto |\Omega|^{-(m+1)/2}, \tag{5.2}$$

where $m = 4$. The indicator function part of the prior restricts positive prior support to the region, $\mathscr{R}$, that satisfies monotonicity conditions for all inputs, the good, and the bad. These conditions have to be evaluated at a particular point in the data set. Due to potential measurement errors, we do not require monotonicity at 100% of our data points. Instead, we define monotonicity as satisfied when the percent of the data points that meet their required monotonicity conditions is 85% for the bad and 97% for the inputs and outputs. However, the actual average number of data points satisfying monotonicity was higher—99% for the inputs and outputs and 92% for the bad.

Write this complete prior as

$$p(\gamma, \Omega) \propto \mathrm{MVN}(\gamma_o, \Psi_o)\mathrm{IW}(\Omega)\mathrm{I}(\gamma, \mathscr{R}), \tag{5.3}$$

where γ_o is the vector of prior means on the parameters in γ, Ψ_o is the prior variance–covariance matrix on the same parameters, and $\mathrm{I}(\gamma, \mathscr{R})$ represents the indicator function that equals one when the restrictions are satisfied and zero otherwise. Zellner (1998) points out the (well-known to Bayesians) danger of

[15] All estimated structural coefficients are in the range of $(-2, 2)$, so this prior variance is very large relative to the magnitude of the parameters.

instrumental variable estimators being bimodal and that imposing sign restrictions can often lead to unimodality of the LIP density.

Given a LIL in the form of (4.10) and the prior in (5.3), the LIP takes the form of

$$p(\gamma, \Omega|D) = c_o|\Omega|^{-(n-k+m+1)/2} \exp[-\tfrac{1}{2}(\gamma - \gamma_p)' \Psi_p^{-1}(\gamma - \gamma_p) - \tfrac{1}{2}\mathrm{tr}(\Xi\Omega^{-1})]\mathrm{I}(\gamma, \mathscr{R}), \tag{5.4}$$

where

$$\gamma_p = \Psi_p(\Psi_o^{-1}\gamma_o + \Psi_m^{-1}\gamma_m), \tag{5.5}$$

$$\Psi_p = (\Psi_o^{-1} + \Psi_m^{-1})^{-1} \tag{5.6}$$

and Ψ_m is the standard GMM covariance matrix of γ_m, which is the standard GMM estimator of γ. The LIP distribution in (5.4) is a truncated version of the standard multivariate normal-inverted Wishart distribution common in Bayesian econometrics.

5.2.2. *Choice of instruments*

Since input and output quantities in all distance function specifications may be endogenous, we examine identification issues using the Hansen (1982) J test. We tested a variety of instrument sets in a standard GMM framework, which included various subsets of outputs and inputs. Interestingly, the data rejected any set of moment conditions that restricted either the outputs or inputs to be exogenous. However, we found support for the use of firm dummies, time period dummies, the interaction of continuous time and firm dummies, the interaction of continuous time-squared and firm dummies, and the interaction of continuous time-cubed and firm dummies. Ultimately, we chose this set of moment conditions as the one that generated the largest p-value for the J test statistic. We also confirm that the instruments are highly correlated with the regressors, by obtaining high R^2 values for the regression of each explanatory variable from (3.1) on our instrument set.[16]

5.2.3. *Analyzing the posterior density*

The most straightforward way to analyze the full joint LIP distribution is to use Gibbs sampling, particularly as we are interested in obtaining posterior measures of a number of nonlinear functions of various elements from within the γ vector. The Gibbs sampler begins with initial values for γ and Ω (we use GMM estimates for this purpose). Then a series of draws are made from the specified conditional posterior distributions for Ω and γ, each conditioned on the most recent random draw of the other. Such a process converges to a random sample from the full joint posterior distribution as in Chib (1995).

A slight complication relative to simple Gibbs sampling is created by our need to truncate the multivariate normal to remain in the area of nonzero prior support. Because the truncation region is highly complex, we chose to simply draw from the full MVN distribution and then discard draws which fall outside the region $\mathscr{R}$ (i.e.,

[16]One might object to the pre-testing we engage in to find the optimal set of instruments. Alternatively, we could do our pre-testing on a training sample and report our results for the remainder of the data.

violate the economic theory). This is essentially an accept-reject algorithm for drawing from a truncated distribution, and still leads to convergence of the draws as shown in Tierney (1994). The first 500 draws were discarded to remove dependence on the initial conditions. We then continued drawing 10,000 more parameter vectors for computation of the posterior distribution. Computation of the posterior standard deviations proved this number of draws to be sufficient. To test convergence, the posterior means were compared to those of other runs of the Gibbs sampler and to subsamples of the 10,000 draws from the run reported here; because these multiple parallel runs and subsamples produced very similar empirical results, we can conclude that our Gibbs sampler has converged. Posterior means are computed as the simple average of the Gibbs draws (or a function of the parameters from each draw), while posterior medians are defined as the median value of a particular parameter or function of parameters from all the draws.

5.3. Results

The results for EC, TC, and PC are presented for both posterior means and medians in Table 3. All results are presented on an annualized basis even though they are measured over five year intervals. Aggregation from firm to industry level is based on weighted averages using electrical generation (output) as the weights. The results show that EC has been negative over the entire time period of our data sample, implying that firm technical efficiencies have been declining over the sample

Table 3
Estimates of EC, TC, and PC

Period	Measure	EC	TC	PC
80–85	Posterior mean	−0.0368	0.1115	0.0747
	Std. dev. of mean	(0.0001)	(0.0003)	(0.0002)
	Posterior median	−0.0371	0.1111	0.0746
	GMM (non-Bayesian)			
	With bads	−0.0411	0.0793	0.0382
	Without bads	−0.0311	0.0686	0.0373
85–90	Posterior mean	−0.0259	0.0915	0.0656
	Std. dev. of mean	(0.0000)	(0.0003)	(0.0002)
	Posterior median	−0.0266	0.0900	0.0636
	GMM (non-Bayesian)			
	With bads	−0.0290	0.0714	0.0425
	Without bads	−0.0248	0.0593	0.0346
90–95	Posterior mean	−0.0183	0.0698	0.0514
	Std. dev. of mean	(0.0000)	(0.0004)	(0.0004)
	Posterior median	−0.0187	0.0685	0.0496
	GMM (non-Bayesian)			
	With bads	−0.0206	0.0622	0.0416
	Without bads	−0.0195	0.0496	0.0301

period. The 2–4% annualized decline (slowing over time) is highly statistically significant given the posterior standard deviations are on the order of 0.01%. This result is not surprising, given the rapid changes in production technology, regulatory structure, and operating environment during this period, and is consistent with one firm pushing the frontier of technology outward, while the other firms are falling behind over time. It is also consistent with estimates of EC computed by Atkinson et al. (2003), who examined a panel from 1961 to 1992 for these same firms, albeit ignoring SO_2 emissions.

TC has been strongly positive over this period, meaning that the technology set has been shifting rapidly outward. From 1980 to 1985, TC was most rapid at 11% per year and then slowed, albeit to a still very rapid 8% average over the remaining 10 years of our data period. While this rate of growth is higher than found by Callan (1991), his data set ends in 1984 and he ignores bads. Our computed rate of TC parallels that found in the technology sector over the 1985–1992 time period. Also, it is important to note that without crediting the utilities for the reductions in emissions, TC estimates would likely have been much lower, with the (non-Bayesian) GMM estimates averaging about 7% over the whole sample and less than 5% from 1990 to 1995, when bads are ignored.

We compute PC as the sum of EC and TC. Since EC was negative and substantial over time, PC was considerably lower than TC. Further, over time, PC falls because the decline in TC was only partially offset by the improvement in EC. From 1990 to 1995, the posterior mean of PC was about 5%, while the non-Bayesian mean ignoring bads was about 3%. For all three measures (EC, TC, and PC) the posterior means and medians were highly similar and displayed no consistent skewness to either the right or left; thus the posterior distributions of these measures appear reasonably symmetric.

Finally, we compute the ratio of the component of TC in (3.9) that involves the quantity of the bad to total TC, as a measure of the percent of TC that is due to crediting the firm for reduction of the bad. On average this effect is 35%. Thus, failing to credit the utilities for their efforts aimed at reducing SO_2 emissions would have understated their true technological progress by 35%, an enormous under-estimate.[17]

Estimated TEs for all 43 firms in the sample are displayed in Table 4. While a detailed analysis for all 43 firms would be excessive, we more closely examine the results for the most and least efficient firms in our sample. The least efficient firm is Alabama Power with a posterior mean TE of 0.2811 (posterior standard deviation of the mean, 0.0008) and posterior median of 0.2718. A symmetric (not shortest) 90% posterior density region for Alabama Power's TE ranges from 0.1603 to 0.4322. The most efficient firm is Rochester (NY) Electric with a posterior mean TE of 0.9090 (posterior standard deviation of the mean, 0.0011) and posterior median of 0.9487. Rochester Electric's 90% posterior density region spans from 0.6916 to 1.0000.

[17]This result is not directly comparable to the estimated 44% loss in PC that Gollop and Roberts (1983) reported due to utilities' efforts to reduce emissions. However, our result appears to be in general agreement with this finding.

Table 4
Time-persistent technical efficiency score by firm

Utility	Classical Tech. eff. score	LIBSE Tech. eff. score
1	0.322312	0.281052
2	0.441017	0.440502
3	0.526488	0.455024
4	0.391878	0.538803
5	0.567188	0.601936
6	0.516249	0.511263
7	0.633104	0.601744
8	0.559161	0.497565
9	0.409722	0.358503
10	0.408242	0.337617
11	0.328289	0.310279
12	0.509245	0.440383
13	0.475729	0.412425
14	0.736671	0.586571
15	0.781180	0.716694
16	0.518828	0.422770
17	0.511422	0.447987
18	0.600381	0.536532
19	0.850217	0.850039
20	0.486643	0.515796
21	0.474907	0.553449
22	0.403431	0.488251
23	0.579200	0.667767
24	0.705174	0.769460
25	0.710635	0.612149
26	0.863582	0.768743
27	0.367026	0.442055
28	0.636714	0.686957
29	0.810497	0.825872
30	0.437101	0.573383
31	1.000000	0.909013
32	0.502580	0.513394
33	0.463269	0.501721
34	0.564591	0.490177
35	0.502142	0.494406
36	0.554549	0.593038
37	0.656814	0.617271
38	0.603356	0.657146
39	0.598090	0.540113
40	0.698688	0.612434
41	0.459108	0.499138
42	0.558731	0.539088
43	0.539753	0.561609
Wtd. avg.	0.564277	0.553187

Table 5
Shadow price of emissions ($/ton)

	1980	1985	1990	1995
Posterior mean	−395.28	−1871.70	−556.83	−486.70
Std. dev. of mean	(27.18)	(16.62)	(2.82)	(2.91)
Posterior median	−102.93	−1629.76	−540.95	−464.68
GMM (non-Bayesian)	−83.00	−1188.88	−17.47	279.17

In order to estimate shadow prices for SO_2 emissions (for which we have incomplete actual prices), we proceed as follows. The shadow price for SO_2 emissions is given in (3.10) (obtained using the envelope theorem applied to the Lagrangian associated with (2.6)). In order to call this price a shadow price, we have assumed that $\mathbf{x}$ is the cost-minimizing solution. To compute the shadow price we need $C([\mathbf{y},\mathbf{b}],\mathbf{p},t)$ and $\nabla_b D_i([\mathbf{y},\mathbf{b}],\mathbf{x}([\mathbf{y},\mathbf{b}],\mathbf{p},t),t)$. However $C([\mathbf{y},\mathbf{b}],\mathbf{p},t)$ is clearly a function of the shadow price vector we desire (shadow costs depend on shadow prices). We resolve this dilemma by assuming that actual prices equal shadow prices for one other input. Assuming that labor is this input, its shadow price is given in (2.7) (derived by taking the first-order condition associated with (2.6)). By taking the ratio of the shadow of the bad to the actual price of labor, $C([\mathbf{y},\mathbf{b}],\mathbf{p},t)$ cancels and we solve for the estimated shadow price of the bad in terms of ratios of estimated partial derivatives and the actual price of labor. We evaluate this expression using the posterior means as estimates of unknown parameters. Note that the shadow price of SO_2 emissions (the price of a bad) is negative, while the price of an SO_2 emissions permit (the price of a good) is positive. Annual aggregates are presented in Table 5.

The Bayesian procedure yields estimated shadow prices that are negative for all years and are very plausible from an economic standpoint. The largest absolute value of the shadow price in 1985 coincides with Title I of the Clean Air Act, which forced Western utilities to invest in expensive scrubbers to reduce the bad. The lower absolute shadow values in 1990 and 1995 likely reflect the technological progress and the switching to low-sulfur Western coal during this period that lowered marginal control costs. Further, beginning in 1995 the U.S. Environmental Protection Agency began a program of tradable emissions permits for SO_2 among electric utilities. Although our data ends just as this program began, the estimated shadow price for 1995, −$487/ton according to the posterior mean or −$465/ton by the posterior median, is more accurate than many other estimates of SO_2 permit prices, although somewhat above actually observed permit prices. For 1995–2000, traded permits generally ranged in price from $100/ton to $200/ton.[18] Also note that for our data, the non-Bayesian procedure yields an estimated shadow price of $279/ton for 1995

[18] In comparison to our estimates, Coggins and Swinton (1996) estimated a shadow price of −$292/ton for a set of plants in Wisconsin; Swinton (1998) estimated values of −$459 in Illinois, −$1298 in Minnesota, and −$33 in Wisconsin; Lee et al. (2002) estimated −$3107/ton; and Färe et al. (2004) estimate values from −$76 to −$142 with their stochastic method. See Ellerman et al. (2000) for a detailed discussion of their and other estimates of SO_2 marginal control costs.

(Table 5), which has the incorrect sign. This partial validation of the Bayesian estimates of shadow prices suggests that the model has performed well, at least compared to other studies regarding SO_2 permit prices, and is likely accurately capturing many of the features of this technology including measures of PC, TC, and EC.

6. Summary and conclusions

Econometric measurement of PC when the firm is controlling bads has typically been carried out using a cost function approach. Inputs are disaggregated into one component that produces bads and another that produces goods. The Divisia index, which assumes constant returns to scale, measures the marginal effect of the former component on costs.

We have avoided this assumption and the pitfall recognized by Pittman (1983) by estimating an input distance function. By treating the bad as an exogenous technology shifter, we also avoided the arbitrary specifications required to transform a bad into a good and the linear homogeneity assumption of the bad with respect to inputs if the bad is scaled proportionally with inputs.

In our econometric estimation we incorporated restrictions from economic theory in our prior distribution, restricting posterior support to parameter vectors that satisfied theoretical conditions such as monotonicity. Following Kim (2002) we estimated a Bayesian limited-information system estimator that minimized assumptions in specifying a LIL function. The GMM specification allows us to easily test for endogeneity, but still combines a traditional prior distribution with the LIL to derive a posterior that can be used in ways identical to standard Bayesian inference. This approach is consistent with the BMOM work of Zellner (1997, 1998), Zellner and Tobias (2001), and Zellner and Chen (2001), who derived post-data densities using maximum entropy methods without utilizing a specific likelihood. They advocate combining their post-data density, which is analogous to our LIL, with priors to obtain a posterior density. Further, the Kim (2002) approach extends their work to the general GMM model, which is more convenient for our analysis.

Through Gibbs sampling we generalized the work of Kim (2002) to incorporate an unknown covariance matrix. Our results showed substantial but declining levels of PC and TC for our sample period, and negative EC over the entire 15 years of our panel data. A considerable portion of PC and TC can be attributed to reduction of the bad. Non-Bayesian estimates of PC and TC which ignore the bad are substantially lower than Bayesian ones which credit the firm for reducing the bad. Finally, shadow values for the bad (SO_2) were estimated which, although higher than actual market prices for emission permits, are clearly more accurate than non-Bayesian estimates and are consistent with other standard econometric estimates of the marginal cost of SO_2 control. This helps to validate the model results and suggests the accuracy of the productivity and efficiency estimates. A logical application of our LIBSE methodology was to the estimation of technical and allocative efficiency modeled as fixed effects in a general panel data context, where we obtained a consistent estimator in the presence of autocorrelation and heteroskedasticity of unknown form.

Acknowledgements

The authors wish to thank Rolf Färe, Chris Cornwell, Gary Koop, Arnold Zellner, Robert Chambers, and audience members at UC Davis, Arizona, Auburn, the 2000 American Agricultural Economics Association Meeting, the 2001 ISBA North American Regional Meeting in Laguna Beach, and the conference that led to this special issue for helpful comments.

References

Atkinson, S., Cornwell, C., 1998. Estimating radial measures of productivity growth: frontier versus non-frontier approaches. Journal of Productivity Analysis 10, 35–46.

Atkinson, S., Cornwell, C., Honerkamp, O., 2003. Measuring and decomposing productivity change: stochastic distance function estimation vs. DEA. Journal of Business and Economic Statistics 84, 284–294.

Baltagi, B.H., Griffin, J.M., 1988. A general index of technical change. Journal of Political Economy 96, 20–41.

Barbera, A., McConnell, V., 1990. The impact of environmental regulations on industry productivity: direct and indirect effects. Journal of Environmental Economics and Management 18, 50–65.

Callan, S.J., 1991. The sensitivity of productivity growth measures to alternative structural and behavioral assumptions: an application to electric utilities 1951–1984. Journal of Business and Economic Statistics 9, 207–213.

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

Chambers, R.G., Chung, Y., Färe, R., 1996. Benefit and distance functions. Journal of Economic Theory 70, 407–419.

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

Chib, S., 1995. Marginal likelihood from the Gibbs output. Journal of the American Statistical Association 90, 1313–1321.

Coggins, J.S., Swinton, J.R., 1996. The price of pollution: a dual approach to valuing SO_2 allowances. Journal of Environmental Economics and Management 30, 58–72.

Conrad, K., Wastl, D., 1995. The impact of environmental regulation on productivity in German industries. Empirical Economics 20, 615–633.

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

Ellerman, D.A., Joskow, P.L., Schmalensee, R., Montero, J.-P., Bailey, E.M., 2000. Markets for Clean Air: The U.S. Acid Rain Program. Cambridge University Press, Cambridge.

Färe, R., Grosskopf, S., Lovell, C., Pasurka, C., 1989. Multilateral productivity comparisons when some outputs are undesirable: a nonparametric approach. Review of Economics and Statistics 71, 90–98.

Färe, R., Primont, D., 1995. Multi-output Production and Duality: Theory and Applications. Kluwer Academic Publishers, Boston.

Färe, R., Grosskopf, S., Chung, Y., 1997. Productivity and undesirable outputs: a directional distance function approach. Journal of Environmental Management 51, 229–240.

Färe, R., Grosskopf, S., Noh, D.-W., Weber, W., 2004. Characteristics of a polluting technology: theory and practice. Journal of Econometrics, this issue.

Fernandez, C., Koop, G., Steel, M., 2000. Multiple-output production with undesirable outputs. Working paper, Department of Economics, University of Edinburgh.

Gollop, F., Roberts, M., 1983. Environmental regulations and productivity growth: the case of fossil-fueled electric power generation. Journal of Political Economy 91, 655–674.

Green, E., Strawderman, W., 1996. A Bayesian growth and yield model for slash pine plantations. Journal of Applied Statistics 23, 285–299.

Hansen, L.P., 1982. Large sample properties of generalized method of moments estimation. Econometrica 50, 1029–1054.

Kim, J.Y., 2002. Limited information likelihood and Bayesian analysis. Journal of Econometrics 107, 175–193.

Koop, G., 1998. Carbon dioxide emissions and economic growth: a structural approach. Journal of Applied Statistics 25, 489–515.

Koop, G., 2001. Cross-sectoral patterns of efficiency and technical change in manufacturing. International Economic Review 42, 73–104.

LaFrance, J., 1999. Inferring the nutrient content of food with prior information. American Journal of Agricultural Economics 81, 728–734.

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, 365–375.

Nelson, R.A., 1984. Regulation, capital vintage, and technical change in the electric utility industry. Review of Economics and Statistics 66, 59–69.

Pittman, R., 1983. Multilateral productivity comparisons with undesirable outputs. The Economic Journal 93, 883–891.

Reinhard, S., Thijssen, G., 1998. Resource use efficiency of Dutch dairy farms: a parametric distance function approach. Working paper, Wageningen Agricultural University.

Swinton, J.R., 1998. At what cost do we reduce pollution? Shadow prices of SO_2 emissions. The Energy Journal 19, 63–83.

Tierney, L., 1994. Markov chains for exploring posterior distributions (with discussion). Annals of Statistics 22, 1701–1762.

U.S. Federal Power Commission. Statistics of Privately Owned Electric Utilities in the U.S.—Classes A and B. US Government Printing Office, Washington, DC, annually.

U.S. Federal Power Commission. Steam Electric Plant Construction Cost and Annual Production Expenses. US Government Printing Office, Washington, DC, annually.

Yaisawarng, S., Klein, J., 1994. The effects of sulfur dioxide controls on productivity change in the U.S. electric power industry. Review of Economics and Statistics 76, 447–460.

Zellner, A., 1971. An Introduction to Bayesian Inference in Econometrics. Wiley, New York.

Zellner, A., 1978. Estimation of population means and regression coefficients including structural coefficients: a minimum expected loss (MELO) approach. Journal of Econometrics 8, 127–158.

Zellner, A., 1997. Bayesian method of moments (BMOM) analysis of mean and regression models. In: Bayesian Analysis in Econometrics and Statistics: The Zellner View and Papers. Edward Elgar, Cheltenham.

Zellner, A., 1998. The finite sample properties of simultaneous equations' estimates and estimators: Bayesian and non-Bayesian approaches. Journal of Econometrics 83, 185–212.

Zellner, A., Chen, B., 2001. Bayesian modeling of economies and data requirements. Macroeconomic Dynamics 5, 673–700.

Zellner, A., Highfield, R.A., 1988. Calculation of maximum entropy distributions and approximation of marginal posterior distributions. Journal of Econometrics 37, 195–209.

Zellner, A., Tobias, J., 2001. Further results on Bayesian method of moments analysis of the multiple regression model. International Economic Review 42, 121–140.