MULTIPLE COMPARISONS WITH THE BEST: BAYESIAN PRECISION MEASURES OF EFFICIENCY RANKINGS

Scott E. Atkinson
Department of Economics
University of Georgia
Athens, GA 30602
email: atknsn@terry.uga.edu
FAX: 706-542-3376

and

Jeffrey H. Dorfman
Department of Agricultural & Applied Economics
University of Georgia
Athens, GA 30602
email: jdorfman@agecon.uga.edu

May 6, 2002

* Paper to be presented at the American Agricultural Economics Association meetings in Long Beach, July 2002. The authors wish to thank John Geweke, Rolf Färe, Chris Cornwell, Robert Town, and audience members at UC Davis, Arizona, the 2001 Canadian Economics Association Meetings, and the 2001 ISBA North American Regional Meeting in Laguna Beach for helpful comments on this or related papers.

Abstract

A large literature exists on measuring the allocative and technical efficiency of a set of firms. A segment of this literature uses data envelopment analysis (DEA), creating relative efficiency rankings that are nonstochastic and thus cannot be evaluated according to the precision of the rankings. A parallel literature uses econometric techniques to estimate stochastic production frontiers or distance functions, providing at least the possibility of computing the precision of the resulting efficiency rankings. Recently, Horrace and Schmidt (2000) have applied sampling theoretic statistical techniques known as multiple comparisons with control (MCC) and multiple comparisons with the best (MCB) to the issue of measuring the precision of efficiency rankings. This paper offers a Bayesian multiple comparison alternative that we argue is simpler to implement, gives the researcher increased flexibility over the type of comparison made, and provides greater, and more intuitive, information content. We demonstrate this method on technical efficiency rankings of a set of U.S. electric generating firms derived within a distance function framework.

Keywords: distance functions, electric utilities, Gibbs sampling, technical efficiency rankings, electric utilities, multiple comparisons with the best.

JEL classification: C11, C32, D24

1. Introduction

Situations abound in which economists, decision makers, and other interested parties desire a ranking of some set according to a chosen metric. Academic departments are ranked according to research output, perceived quality of faculty, and/or reputation. Hospitals are ranked according to mortality rates (often adjusted for severity of the injuries they treat). Firm's are ranked relative to intra-industry competitors on the basis of technical efficiency. In all these situations, in addition to the desired ranking, it would be beneficial to provide information on the precision of the rankings. In laymen's terms, can we truly differentiate the units of observation or are we more accurately perhaps only separating them into groups? In extreme cases, a set of firms might be ranked by efficiency, yet the most and least efficient firms might not truly be distinguishable due to a lack of statistical precision. In such a case, the ranking would be best suppressed.

A huge literature exists on measuring the relative efficiency of a set of firms, in both allocative and technical senses. A segment of this literature uses data envelopment analysis (DEA), creating relative efficiency rankings that are nonstochastic and thus cannot be evaluated according to the precision of the rankings. A parallel literature uses econometric techniques, such as stochastic production frontiers or estimation of distance functions, providing at least the possibility of computing the precision of the resulting efficiency rankings. Recently, Horrace and Schmidt (2000) have applied sampling theoretic statistical techniques known as multiple comparisons with control (MCC) and multiple comparisons with the best (MCB) to the issue of measuring the precision of efficiency rankings. This technique allows researchers and users of such rankings to discover the precision with which certain firms can be ranked above others, along with discovering sets of firms that

are statistically tied with each other even if the point estimates of their relative efficiencies differ.

In this paper we offer a Bayesian alternative that we will argue is simpler to implement, more flexible over possible comparisons, and provides greater, and more intuitive, information content. The Bayesian method easily allows comparisons between single firms, a firm versus a group, or a group versus a group. Further, rather than simply answering the question of "can we differentiate?" with a yes/no (reject/do not reject), the Bayesian method provides an estimated probability in support of the rankings ability to differentiate between the two firms or groups compared. Thus, statements such as "firm A can be ranked as more efficient than firm B with a 92 percent posterior probability" are possible. We demonstrate this method on technical efficiency rankings of a set of U.S. electric generating firms derived within a distance function framework.

The remainder of this paper is organized as follows. In section 2, we review the MCB and MCC approaches pioneered by Horrace and Schmidt for the purposes of efficiency rankings. In section 3, we introduce the Bayesian approach and discuss differences and potential advantages to the Bayesian methodology. Section 4 discusses the model, the data, an overview of the derivation of our efficiency rankings, and the results of our empirical application. In particular, we focus on the results produced by the Bayesian multiple comparison approach and contrast them with the original MCC and MCB approaches. Conclusions follow in section 5.

2. MCB and MCC Approaches to Testing Efficiency Rankings

Horrace and Schmidt (2000) pioneered the use of MCB and MCC in creating statistical confidence intervals for use with comparisons of multiple firm efficiency scores. Their

procedures allow some hypothesis tests to be conducted in a sampling theory framework so that researchers can state whether a firm is "significantly" more efficient than some group of firms.

While Horrace and Schmidt (2000) focuses on MCB, MCC seems the more natural application. Referring to efficiency rankings for concreteness, the distinction is that MCC involves comparing the estimated efficiency of a chosen (and fixed) firm to another firm or group of firms while MCB adjusts for the case where the "best" or index firm is unknown. It is clear that once one recognizes the stochastic and imprecise nature of the estimated rankings, one should also realize that the most efficient (best) firm is unknown. However, in most real world application (as opposed to academic ones), it is quite reasonable to use the firm estimated to be best as the index firm and investigate how many of the other firms can be declared statistically less efficient. Choosing this index firm as fixed leads one to the MCC algorithm, so we explain that first.

2.1 The MCC Method

Begin by denoting the estimated measure for each firm i (technical efficiencies in the application to follow) by $\theta_i, i = 1, \ldots, N$. Assume for simplicity that firms were ordered in such a way that θ_N has the largest measure (highest efficiency) and is thus the best, or index, firm against which we wish to compare the others. The MCC method computes a joint confidence interval of a desired probability level for all the differences between individual firm efficiencies and the best. That is, for the vector $[\theta_N - \theta_1, \theta_N - \theta_2, \ldots, \theta_N - \theta_{N-1}]$. When the efficiency estimates are independent, this joint confidence interval can be given by equation (5) from Horrace and Schmidt (2000), rewritten slightly here as

$$(1-\alpha) = \text{Prob}(\hat{\theta}_N - \hat{\theta}_i - h \leq \theta_N - \theta_i \leq \hat{\theta}_N - \hat{\theta}_i + h, \ \ \forall\, i = 1, N-1), \qquad (2.1)$$

where $h = d(2k\hat{\sigma}^2)^{1/2}$, d is the critical value for the joint two-sided confidence interval which has been adjusted to account for the multiple comparisons, and k is the factor of proportionality which scales the identity covariance matrix of $\hat{\theta}$. For details see Horrace and Schmidt (2000), equation (6). Tables of critical values for MCC can be found in Hahn and Hendrickson (1971), *inter alia*. Horrace and Schmidt (2000) also discuss how to extend the results to cases where the efficiency estimates are correlated (the most common case).

Given the joint confidence interval, Horrace and Schmidt (2000) identify all firms that are statistically less efficient than the best firm, along with all the firms that cannot be differentiated from the best. These two groups of firms are simply those for whom the joint confidence intervals, respectively, do not and do include a zero difference at the chosen significance level.

2.2 The MCB Method

The extension from the MCC to the MCB method is that now the best firm is considered unknown, implying that each firm's efficiency needs to be compared to a best firm whose identity is uncertain. Thus in equation (2.1), we would need to replace the fixed index firm θ_N with an unknown best index firm, $\theta_{(N)}$, in Horrace and Schmidt's notation. This somewhat complicates the construction of the joint confidence interval, but the simplified version of the results is that the set of firms which cannot be statistically differentiated from the uncertain best firm are those for which

$$\hat{\theta}_j - \hat{\theta}_i \leq h, \quad \forall\, j \neq i, \tag{2.2}$$

where h is the same as in equation (2.1) and represents one half of the width of the confidence interval. All firms for which the condition in equation (2.2) holds are in the set of possible best firms as defined by Horrace and Schmidt, although technically this is the

set of firms whose estimated efficiency measures are not statistically significantly below the measure of the uncertain most efficient firm. Those firms for which the condition in equation (2.2) does not hold are obviously outside that set and can be said to be statistically less efficient than the best firm.

To extend MCC and MCB to the case in which the estimated TEs are correlated (the common case), the h described here becomes a function of the firms being compared and so is replaced by a comparison-specific h_{ji}. This comparison-specific confidence interval width is computed by multiplying a firm-specific adjusted critical value d_j and a comparison-specific covariance $\hat{\sigma}_{ji}$ that replaces $2k\hat{\sigma}^2$ in the formula for h from section 2.2.

3. A Bayesian Approach to Measuring the Precision of Efficiency Rankings

In contrast to the sampling theory approach outlined above, we show in this section that a Bayesian approach can be taken using the empirical results that arise naturally from the Markov Chain Monte Carlo (MCMC) algorithm employed to derive the numerical estimates in our application and from any other numerical Bayesian estimation technique. This Bayesian Multiple Comparison (BMC) methodology provides exact (posterior) probability levels for each comparison statement to be evaluated. Thus, rather than simply stating that "firm A is (not) significantly more efficient than all firms in group 4 using a 5 percent significance level," we can make statements along the lines of "there is an estimated 97 percent probability that firm A is more efficient than all firms in group 4" and "there is only a 15 percent probability that firm A is more efficient than all firms in group 4." These statements contain much more information and a much higher degree of specificity than the ones developed using the MCC/MCB framework.

The BMC is simple to implement with the parameter draws, generated in our application by a Gibbs sampling algorithm, which we use to compute posterior estimates of the

unknown parameters and the technical efficiency scores. As will be detailed precisely in the next section, numerical Bayesian techniques rely on random draws from throughout the parameter space to generate approximate values for parameters of interest, functions of the model parameters (such as efficiency measures), precision measures, and probability levels in support of hypotheses of interest. The preciseness of the numerical approximation is controlled by the choice of the Bayesian numerical technique and the number of parameter draws generated, so researchers can obtain any desired level of precision.

Reserving the discussion of exactly how to get a set of such draws for the next section, for now it suffices to establish that given a set of random draws from the posterior density function of a vector of parameters θ one can estimate the posterior mean of a function of interest, say $g(\theta)$, by the arithmetic mean of the draws. See for example, Tierney (1994). The technical efficiencies which researchers want to compare are just such a function of interest and can be expressed as a function of the randomly drawn parameter vector.

Each Gibbs draw is used to compute TE scores for each firm, denoted by $TE_i^{(b)}$ for firm i and draw b. In addition to using these draws to find posterior means, medians, standard deviations, they can be compared across firms. To estimate the probability that firm i is more efficient than firm j, we count the number of draws for which firm i's TE score is greater. Formally, for a set of B draws on the TE scores,

$$\text{prob}(\text{TE}_i > \text{TE}_j) = B^{-1} \sum_{b=1}^{B} H(\text{TE}_i^{(b)} > \text{TE}_j^{(b)}), \tag{3.1}$$

where H is a logical operator equal to one when the argument is true and zero otherwise. If one uses a different numerical Bayesian approach that yields draws where weights are needed to arrive at accurate posterior means (such as in importance sampling), then the weights would scale the right-hand side of equation (3.1) above. To estimate probabilities

for multiple comparisons, simply replace one or both of the single TE scores with the sets desired. For example, to compare firm i to a group J, the logical operator would evaluate the truth of $\mathrm{TE}_i^{(b)} > \max\{\mathrm{TE}_j^{(b)}, j \in J\}$.

These probability levels essentially create a Bayesian analog to the MCC procedure, with the advantage of simplicity and greater information content on the strength of support in favor (or against) differentiation between compared firms or groups. However, since the index firm has been fixed, an extension of the above procedure is necessary to generate a Bayesian MCB. While the frequentist idea of joint confidence intervals for differences between TE scores does not translate perfectly into the Bayesian framework, one could create a Bayesian analog. Rather than create a single analog, we choose to list several possible Bayesian MCB-type measures.

Defining J as the set of all firms other than firm i and retaining the above definition for the logical operator H, one can estimate the probability that firm i is the most efficient firm by

$$\mathrm{prob}(\mathrm{TE}_i = \mathrm{TE}_{\max}) = B^{-1} \sum_{b=1}^{B} H(\mathrm{TE}_i^{(b)} > \max\{\mathrm{TE}_j^{(b)}, j \in J\}). \tag{3.2}$$

Given some value δ, chosen perhaps to represent an economically significant difference in TE scores, one can compute the probability that a firm's TE score lies within the specified range of the best:

$$\mathrm{prob}(\mathrm{TE}_{\max} - \mathrm{TE}_i \leq \delta) = B^{-1} \sum_{b=1}^{B} H(\max\{\mathrm{TE}_j^{(b)}, j \in J\} - \mathrm{TE}_i^{(b)} \leq \delta). \tag{3.3}$$

This equation can clearly be used to create an analog to the MCB procedure's set S of firm's in contention to be the best. Simply allow i to vary for firms $i = 1, \ldots, N$ and

place all firms in the set S that yield probabilities of greater than some prespecified level (1-α) for being within a distance δ of the best TE score. Because of the small differences between Bayesian MCC and MCB analogs, we will simply refer to the BMC procedure without using an additional C or B designation for control or best.

Within a numerical Bayesian estimation framework, whether dealing with simple Monte Carlo integration, MCMC approaches such as Gibbs sampling, or even importance sampling, one can always estimate the probability of a ranking being accurate (or correct) by simple evaluation of the frequency of the ranking occurring within the large set of random parameter draws employed in the numerical integration. For more discussion of the foundations of numerical Bayesian methods, see Geweke (1999).

Horrace and Schmidt (2000) do not perform comparisons between groups or of a single firm versus another single firm or subgroup. While the sampling theory MCC and MCB approaches can be extended to accomplish the same tasks just introduced with the Bayesian approach, to accomplish these different types of comparisons the statistical foundation of their procedure must be resolved to yield the correct critical values for each such comparison. Because the adjustment for the multiple comparisons is conditional on the nature and number of such comparisons made, the MCC and MCB algorithms must be adjusted whenever the format of the multiple comparisons changes. Thus, while there is no theoretical barrier to stop MCC and MCB approaches from performing the same sorts of comparisons as our BMC approach, the task is daunting until more work is done to develop user-friendly software.

4. Empirical Application

We apply our methodology to a panel of U.S. electric utilities observed at five-year intervals from 1980–1995. There are two outputs: the quantity of electric power generated

(a good output) and the quantity of SO_2 emissions (a bad output which locally has a direct negative effect on health and welfare and regionally can lead to acid rain). Three inputs are applied to produce these outputs: 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 and since electricity is currently in short supply in the State of California, where State Implementation Plans are very strict. 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 for these utilities. Proper crediting for reduction of this bad is essential to obtain unbiased estimates of efficiency levels. It also can provide insights into what the tradeoff has been between emissions and output.

4.1 The Model

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 good outputs 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

$$S(\mathbf{x}, \mathbf{y}, t) = \{(\mathbf{x}, \mathbf{y}) : \mathbf{x} \,\text{can produce}\, \mathbf{y} \,\text{at time t}\}, \tag{4.1}$$

where $t = 1, \ldots, T$ is time.

This application, however, has a bad output (air pollution) that must be accounted for to accurately measure the technical efficiency of the various utilities. Ignoring the bad would allow a firm to look more efficient by ignoring the environment, while a firm that

spent effort on mitigating its pollution would be unrewarded and measured as relatively inefficient.

Symmetric treatment of bads (denoted by a vector $\mathbf{b}$) and goods using an input distance function is legitimate and can be specified as

$$D_i([\mathbf{y},\mathbf{b}],\mathbf{x},t) = \sup_{\lambda}\{\lambda : ([\mathbf{y},\mathbf{b}],\mathbf{x}/\lambda) \in S(\mathbf{x},\mathbf{b},\mathbf{y},t)\}. \tag{4.2}$$

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 (4.2) is dual to the cost function, we can write

$$C_i([\mathbf{y},\mathbf{b}],\mathbf{p},t) = \min_{\mathbf{x}}\{\mathbf{px} : D_i([\mathbf{y},\mathbf{b}],\mathbf{x},t) \geq 1\}, \tag{4.3}$$

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 unit cost function if costs are minimized. 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

$$\mathbf{p} = C([\mathbf{y},\mathbf{b}],\mathbf{p},t)\nabla_x D_i([\mathbf{y},\mathbf{b}],\mathbf{x},t). \tag{4.4}$$

Equivalently, the bads can be treated as exogenous shifters of the technology set, similar to a time trend or state of technology variable. The intuition 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 reducing output levels 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, equation (4.2) can be written as

$$D_i(\mathbf{y}, \mathbf{x}, t|\mathbf{b}) = \sup_{\lambda}\{\lambda : (\mathbf{x}/\lambda, \mathbf{y}|\mathbf{b}) \in S(\mathbf{x}, \mathbf{y}, t|\mathbf{b})\}. \tag{4.5}$$

The appropriate monotonicity condition for the bad in the context of the input distance function can be derived as follows. Assuming a single bad, we compute the partial total differential of equation (4.5) evaluated on the frontier at a fixed time [implying $D_i(\mathbf{y}, \mathbf{x}, t|\mathbf{b}) = 1$ and $dt = 0$] to obtain

$$dD_i = \sum \frac{\partial D_i}{\partial y_m} dy_m + \sum \frac{\partial D_i}{\partial x_n} dx_n + \frac{\partial D_i}{\partial b} db = 0. \tag{4.6}$$

Using the properties that the input distance function is monotonically nondecreasing in inputs ($\frac{\partial D_i}{\partial x_n} \geq 0$) and monotonically nonincreasing in outputs ($\frac{\partial D_i}{\partial y_m} \leq 0$), and setting $dy_m = 0, \forall m$, in order to keep the firm on the input distance frontier, we obtain

$$\frac{\partial D_i}{\partial b} = -\sum \frac{\partial D_i}{\partial x_n} \frac{dx_n}{db}. \tag{4.7}$$

As in Pittman (1983), with constant desirable output and technology, bads can only be reduced through increased input usage. This implies that $\frac{dx_n}{db} \leq 0$, which combined with the nonnegativity property for inputs, $\frac{\partial D_i}{\partial x_n} \geq 0$, yields

$$\frac{\partial D_i}{\partial b} \geq 0. \tag{4.8}$$

As a flexible approximation to the true distance function in (4.5) , 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

$$
\begin{aligned}
0 = {} & \gamma_0 + \sum_m \gamma_m \ln y_{mft} + \sum_z \gamma_z \ln b_{zft} \\
& + \sum_n \gamma_n \ln x_{nft} + \gamma_{t1} t + (1/2)\gamma_{t2} t^2 \\
& + (1/2) \sum_m \sum_{m'} \gamma_{mm'} \ln y_{mft} \ln y_{m'ft} + (1/2) \sum_z \sum_{z'} \gamma_{zz'} \ln b_{zft} \ln b_{z'ft} \\
& + (1/2) \sum_n \sum_{n'} \gamma_{nn'} \ln x_{nft} \ln x_{n'ft} + \sum_m \sum_n \gamma_{mn} \ln y_{mft} \ln x_{nft} \\
& + \sum_z \sum_n \gamma_{zn} \ln b_{zft} \ln x_{nft} + \sum_z \sum_m \gamma_{zm} \ln b_{zft} \ln y_{mft} \\
& + \sum_m \gamma_{mt} \ln y_{mft} t + \sum_z \gamma_{zt} \ln b_{zft} t \\
& + \sum_n \gamma_{nt} \ln x_{nft} t + \ln h(\epsilon_{ft}),
\end{aligned}
\tag{4.9}
$$

where

$$
h(\epsilon_{ft}) = \exp(v_{ft} - u_{ft}), \tag{4.10}
$$

so that $\ln h(\epsilon_{ft})$ is an additive error with a one-sided component, u_{ft}, and a standard noise component, v_{ft}, with zero mean.[1]

In principle, the u_{ft} can be treated as fixed or random, but the choice between the two entails a tradeoff. With the fixed effects approach, identification is potentially difficult, since the number of parameters increases with the number of firms, F. 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, Schmidt, and Sickles (1990), we choose a specification of the form,

$$
u_{ft} = \sum_{q=0}^{Q} \beta_{fq}\, d_f\, t^q, \qquad f = 1, \ldots, F, \tag{4.11}
$$

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

where t is a trend and d_f is a dummy variable equal to one for firm f and zero for the other firms.[2] With a fixed effects approach, the β_{fq} are firm-specific parameters to be estimated. This avoids the distributional and exogeneity assumptions that would otherwise be required in a random effects setup. Thus, the estimated equation is obtained by substituting (4.11) into (4.10), which in turn is substituted into (4.9), so that the β_{fq} are fit directly with the other parameters.

In the application that follows, we undertake a Bayesian method of moments estimation based partially on the moment conditions $E(v_{ft} \,|\, \mathbf{z}_{ft}) = 0$, where $\mathbf{z}_{ft}$ is a vector of instruments. In distance function applications, it is highly unlikely that $(\ln y_{ft}, \ln x_{ft})$ will be uncorrelated with v_{ft}, thus pointing to the need for an instrumental variables approach.

Since we do not impose one-sidedness (non-negativity) on the u_{ft} in estimation, we need to do so after estimation, by adding and subtracting from the fitted model $\hat{u}_t = \min_f(\hat{u}_{ft})$, which defines the frontier intercept. With $\ln \hat{D}(\mathbf{y}, \mathbf{x}, t)$ representing the estimated translog portion of (4.9) (i.e., those terms other than $h(\epsilon_{ft})$), adding and subtracting $\hat{u}_t$ yields

$$0 = \ln \hat{D}_i(\mathbf{y}, \mathbf{x}, t) + \hat{v}_{ft} - \hat{u}_{ft} + \hat{u}_t - \hat{u}_t = \ln \hat{D}_i^*(\mathbf{y}, \mathbf{x}, t) + \hat{v}_{ft} - \hat{u}_{ft}^*, \tag{4.12}$$

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 $\hat{u}_{ft}^* = \hat{u}_{ft} - \hat{u}_t \geq 0$.

Using (4.11), we estimate firm f's level of technical efficiency in period t, TE_{ft}, as

$$TE_{ft} = \exp(-\hat{u}_{ft}^*), \tag{4.13}$$

where our normalization of $\hat{u}_{ft}^*$ guarantees that $0 < TE_{ft} \leq 1$.

[2] An alternative approach by Koop (2001) parameterizes the mean of an exponential technical inefficiency distribution using a vector of variables thought to correlate with firm-specific effects.

Prior to estimation, several sets of parametric restrictions are imposed on (4.9). We impose symmetry, linear homogeneity in input quantities, and constrain β_{fq}, $\forall q$, to equal zero for one firm in order to achieve identification. Symmetry requires that

$$\begin{aligned}
\gamma_{mm'} &= \gamma_{m'm}, \forall\ m, m',\ m \neq m' \\
\gamma_{zz'} &= \gamma_{z'z}, \forall\ z, z',\ z \neq z' \\
\gamma_{nn'} &= \gamma_{n'n}, \forall\ n, n',\ n \neq n'.
\end{aligned} \tag{4.14}$$

In addition, linear homogeneity in input quantities implies

$$\begin{aligned}
\sum_n \gamma_n &= 1, \\
\sum_n \gamma_{nn'} = \sum_{n'} \gamma_{nn'} &= \sum_n \sum_{n'} \gamma_{nn'} = 0, \\
\sum_n \gamma_{mn} &= 0, \forall\ m, \\
\sum_n \gamma_{zn} &= 0, \forall\ z, \text{ and} \\
\sum_n \gamma_{nt} &= 0.
\end{aligned} \tag{4.15}$$

Finally, identification requires that β_{fq}, $\forall q$, must be constrained for one firm in (4.11).

4.2 Data

Our dataset is an updated and refined version of the panel of utilities originally analyzed by Nelson (1984).[3] 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

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

utilities for the years 1980, 1985, 1990 and 1995.[4] 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.* Over the 1980–1995 time period, x_K declined somewhat. More dramatic was the greater than 20 percent reduction in x_L and b. Finally, x_E and y increased moderately.

Data on SO_2 emissions is published on the EPA Acid Rain Website.[5] The primary data is for Clean Air Act Amendment Phase I and Phase II units and was 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,

[4] The primary sources for Nelson's sample are the 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*.

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

our panel is comprised of 43 firms for the years 1980, 1985, 1990, and 1995, for a total of 172 observations.

4.3 Bayesian Estimation Procedure

We estimate the model in a Bayesian generalized method of moments (BGMM) framework. To do so, we must specify priors for the unknown parameters and a set of moment conditions for the data. We can then follow earlier work to find the maximum entropy (maxent) density that is compatible with our prior and moment information. In total, we have a system of four equations: the distance function and three first-order conditions for the three inputs. Our BGMM approach follows and extends Zellner (1998) and Zellner and Tobias (2001) which both present estimates using a Bayesian method of moments (BMOM) approach. Our extensions allow the use of instruments to address the endogeneity inherent in estimation of a distance function, the nonlinear nature of our system of equations, and the incorporation of informative priors on the random parameters while still yielding exact finite sample posterior moments for the parameters of interest (Zellner, 1998). To implement the BGMM algorithm, we combine two sets of moment conditions and a proper prior density, yielding a proper posterior density for the unknown random parameters.[6]

Our estimated distance system consists of (4.9), subject to (4.10) and (4.11), and a set of first order conditions in (4.4), for a total of four equations. As indicated, we impose symmetry and linear homogeneity. An additive iid error term, $w_{ft}^k, k = 1, \ldots, 3$, is appended to each equation in (4.4). After setting $Q = 2$ in (4.11), we test a number of null hypotheses by computing 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 .01 level we fail to reject the null hypothesis that $\beta_{f0} = 0, \forall f$ and subsequently drop the corresponding firm dummies.

[6] For full details of the estimation algorithm, see Atkinson and Dorfman, 2001.

Similarly, we fail to reject the null hypothesis that $\beta_{f2} = 0, \forall f$ and therefore Q is set equal to 1 in (4.11). Further, we set $\beta_{11} = 0$ to achieve identification.

Since input and output quantities in all distance function specifications may be endogenous, we use an instrumental variables approach. To examine identification issues we use the Hansen (1982) J test. We found support for the use of the set of instruments containing 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. This set of moment conditions generated the largest $p-$value for the J test statistic. We also confirm that the instruments are highly correlated with the regressors.

4.3.1 Specification of the Prior

In the specification of the prior, we differ from Zellner (1998) and Zellner and Tobias (2001), by going beyond a maxent prior to a more informative one. The full prior distribution is 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 theoretical restrictions are satisfied.

The structural parameters of the distance function are each given a normal prior distribution with zero mean and variance of 100. This is a very diffuse prior, having virtually no effect on 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 posterior sample density more straightforward to work with when we begin Gibbs sampling. The prior for Σ (the matrix of variances and covariances of the four errors appended to the equations to be estimated) is a standard Jeffreys prior.

The indicator function part of the prior restricts positive prior support to the region, $\mathcal{R}$, that satisfies a set of conditions derived from economic theory. Monotonicity is required 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 85% of the data points meet their required monotonicity conditions.

We can write this prior distribution as the product of its three parts: a multivariate normal for the γ parameters, the Jeffreys prior on Σ, and an indicator function to represent the restrictions from economic theory. Write this as

$$p(\gamma, \Sigma) \propto \text{MVN}(g_o, H_o)|\Sigma^{-1}|^{-5/2}\text{I}(\gamma, \mathcal{R}), \tag{4.16}$$

where g_o is the vector of prior means on the parameters in γ, H_o is the prior variance-covariance matrix on the same parameters, and $I(\gamma, \mathcal{R})$ represents the indicator function that equals one when the restrictions are satisfied and zero otherwise.

4.3.2 The Posterior Density

Following Zellner (1998) and Zellner and Tobias (2000), a maxent framework is used to yield the least informative posterior density that is still proper and consistent with the prior in (4.16) and the first and second moment conditions specified by our instrumental variables, generalized method of moment approach. This joint density is a truncated version of the standard multivariate normal-inverted Wishart distribution common in Bayesian econometrics,

$$p(\gamma, \Sigma|\text{data}) \sim \text{MVN}{-}\text{IW}(g_p, H_p)\text{I}(\gamma, \mathcal{R}), \tag{4.17}$$

where MVN is a multivariate normal density, IW is an inverted Wishart density,

$$H_p^{-1} = H_o^{-1} + H_d^{-1} \tag{4.18}$$

and

$$g_p = H_p[H_o^{-1} g_o + H_d^{-1} g_d], \tag{4.19}$$

where g_d is the conventional GMM estimator of γ and H_d is the conventional GMM estimated covariance matrix of γ.

Because the joint posterior density is complicated to deal with due to the prior restrictions, we use Gibbs sampling to generate draws sequentially from conditional distributions of parameter subsets.[7] In this model, we only need two subsets. First, we can draw the covariances from an inverted Wishart distribution conditional on the previous draw for the γ vector. Then the γ vector can be drawn from a truncated multivariate normal distribution conditional on the drawn value of the Σ matrix. In terms of a "recipe," the Gibbs sampler in our application is comprised of the following steps:

0. Obtain initial value for covariance matrix of errors, $S^{(0)}$, either through conventional GMM estimation as $\hat{\mathbf{e}}'\hat{\mathbf{e}}/\nu$, where the $(4FT \times 1)$ column vector $\mathbf{e} = (v', w_1', w_2', w_3')'$, or after arbitrary choice of all parameters.
1. Draw $\Sigma^{(i)}$ from IW$(S^{(i)}, \nu)$, where $\nu = FT - K$, and K is the number of estimated parameters (Draw system covariance matrix conditional on covariance estimate in 1.) See the Appendix for futher details on this step.
2. Compute $g^{(i)} = \text{GMM}(y, X, Z|\Sigma^{(i)})$, (Compute GMM estimate conditional on $\Sigma^{(i)}$) This requires iterating until convergence using GMM with the covariance of the errors held constant at $\Sigma^{(i)}$.

[7] For a good and simple explanation of Gibbs sampling for the non-Bayesian, see Casella and George (1992).

3. Compute $H^{(i)} = \text{cov}(g^{(i)})$ (Calculate estimated covariance of GMM coefficients)
4. Compute $H_p^{(i)} = [H_o^{-1} + (H^{(i)})^{-1}]^{-1}$ (combining prior variance with data variance to get posterior variance)
5. Compute $g_p^{(i)} = H_p^{(i)}[H_o^{-1}g_o + (H^{(i)})^{-1}g^{(i)}]$ (Combine prior mean of coefficients with moment conditions through the maxent principle to get posterior mean.)
6. Draw $\gamma^{(i)}$ from MVN$(g_p^{(i)}, H_p^{(i)})$ (Draw candidate parameters from a multivariate normal distribution)
7. If $\gamma^{(i)} \in \mathcal{R}$, continue, otherwise go back to step 6 (Satisfy restrictions that impose economic theory)
8. Compute $S^{(i)}$ from the residuals.
9. Return to step 1 conditioning on new values of all parameters.

In the above, GMM represents an operator to compute a standard GMM estimator with four arguments representing the y data, the X data, instruments, and fixed covariance matrix, respectively. To begin this procedure, arbitrary initial values $\gamma^{(0)}$ and $S^{(0)}$ are needed; we use GMM estimates for this purpose. Then steps 1 through 8 are repeated in a loop, each step conditioned on the most recent values of all other parameters and values in the process. Such a process converges to a random sample from the full joint posterior distribution as in Chib (1995). For details on performing MCMC with these and other distributions, see Tanner (1996).

The first 500 draws were discarded to remove dependence on the initial conditions. We then continued drawing 3,000 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 3,000 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.

4.4 Results

Estimated TEs for all 43 firms in the sample are displayed in Table 2. 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 0.2795 (posterior standard deviation of the mean, 0.0015) and posterior median of 0.2705. A symmetric (not shortest) 90% highest posterior density region for Alabama Power's TE ranges from 0.1625 to 0.4306. The most efficient firm is Rochester (NY) Electric with a posterior mean TE of 0.9115 (posterior standard deviation, 0.0020)and posterior median of 0.9563. Rochester Electric's 90% highest posterior density region spans from 0.6888 to 1.0000.

An analysis of the 43 firms' estimated technical efficiencies suggests grouping the firms into four groups. Group 1 (G1) contains the seven firms with the highest posterior mean TEs, all of which have at least a 90% posterior probability of being more efficient than at least 25 other firms. Group 2 (G2) contains the next 14 firms, representing the remainder of the top half of the firms when sorted by posterior mean TE. Group 3 (G3) contains the next 18 firms in this ranking by posterior mean. Finally Group 4 (G4) contains the bottom four firms, the least efficient according to the posterior mean TEs. These firms were placed in Group 4 due to their all having less than a 50% posterior probability of being more

technically efficient than any single firm outside of G4. Our firms are identified by these groupings in Table 2, with the groups identified by the numbers 1 (most efficient) through 4 (least efficient). The posterior mean estimates of each firm's TE are also displayed graphically in Figure 1, sorted from least to most efficient along with their firm IDs and group numbers. One can clearly see differentiation between the most and least efficient groups and the firms in the middle two groups. Visually differentiating between G2 and G3 is more problematic. This visual information motivates us to go beyond the firm by firm analysis briefly mentioned above used in initially categorizing the firms. A use of Bayesian multiple comparisons will allow us to precisely define which firms can be differentiated from each other.

4.4.1 Bayesian Multiple Comparisons

To present TE results obtained using the BMC approach we will use the two firms identified as the most efficient, firm 31 (F31), and the least efficient, firm 1 (F1), and also the groups of firms (designated G1, G2, G3, and G4 for this section). Because we hold the comparison units (firm or group) constant, this is analogous to what Horrace and Schmidt call MCC.[8]

Results of comparing each of the four efficiency groups to the others are presented in Table 3. As can be seen in the table, G4 (the least efficient group) can be differentiated from G2 and G1 with high probability levels, implying that all firms in G4 are almost surely less efficient than all firms in both G1 and G2. However, G3 is more efficient as a group than G4 at a probability support level that would not satisfy many researchers

[8] The numerical Bayesian approach easily adapts to the MCB algorithm of an unknown "best" firm. To compute a probability that firm A is less efficient than the "best" or most efficient firm is likely to result in a probability near 1 given that firm A's efficiency level cannot be greater than that of the most efficient firm in any draw, only equal to it.

(prob. = 0.383). Similarly, we find that G2 rarely dominates G3, suggesting that these groups are not clearly differentiated at any meaningful level of statistical precision.

Moving on to firm-specific comparisons, the most and least efficient firms can be compared to the remainder of their respective groups to examine whether they are clearly identified as best and worst. The posterior probability that F1 is less efficient than the remainder of G4 is 0.641, reflecting a reasonable confidence in this ranking, but not definitive support. The posterior probability that F31 is more efficient than the remainder of G1 is 0.403, indicating that it does not necessarily deserve to be overly singled out as superior, although this probability still greatly exceeds the expected probability if all firms in G1 were equally efficient (recall that the group contains a total of seven firms).

Finally, proceeding to the comparisons of the index firms to the other groups, we begin with F1 again. F1 has a posterior probability of being less efficient than G3 equal to 0.933, of being less efficient than G2 equal to 0.990, and of being less efficient than G1 equal to 0.994. All of these indicate enormous evidence in favor of the precision of the last-place ranking of this firm relative to all firms in the top 3 groupings. F31 has an estimated posterior probability of being more efficient than G4 of 0.996, than G3 of 0.917, and than G2 of 0.834. Thus, F31 appears to be properly ranked above the firms in the other groups.

The BMC comparisons in this section clearly do not exhaust all possible subsets. Our intent is only to convey the richness of the possible types of comparisons that can be easily performed.

4.4.2 Comparing Bayesian to Sampling Theory MCB

The basic sampling theory MCB algorithm of Horrace and Schmidt (2000) provides a set of firms which cannot be statistically differentiated from the uncertain best firm,

as detailed above in section 2.2. When performed on the 43 firms in our sample using a 95% significance level, this set contains 29 firms; thus, only 14 firms are statistically less efficient than the best.

Comparing these results to the values in Table 3 finds reasonable concurrence with some interesting differences. All seven firms in G1 are in the set of possible best firms computed by the MCB approach along with 12 out of 14 firms from G2. The two firms from G2 excluded are F5 and F30, neither of which is near the bottom of the group in terms of posterior mean TE. In fact, F5 is the median firm within G2. Rounding out the set are 10 firms from G3, including F13 and F16 which have the 2 smallest posterior mean TEs within G3. A The set of possibly efficient firms is denoted in Table 2 by a * in the MCB efficient set column.

The inclusion of firms in the sampling theoretic MCB approach's efficient set that are fairly soundly rejected by the Bayesian approach (such as F13 and F16) is somewhat difficult to explain. One possible explanation is that the sampling theory MCB approach produces a conservative joint confidence interval. However, this cannot explain the fact that the MCB approach rejects equality for firms with smaller gaps in point estimates while fails to reject equality for firms with larger estimated efficiency gaps. Examining these four firms further from the Bayesian side is interesting. The two firms from G2 that are excluded from the MCB efficient set, F5 and F30, have Bayesian posterior probabilities of being more efficient than F31 of only 2.2% and 2.8%, respectively, so their exclusion appears to make sense. Yet the two bottom firms from G3 which are included in the MCB efficient set, F13 and F16, both have a Bayesian posterior probability of being more efficient than F31 of only 0.6%. This makes their inclusion while F5 and F30 are excluded even more puzzling.

5. Conclusions

While many researchers have developed and applied methods for estimating the technical efficiency of firms (or other units of observations), less effort has been expended on examining the precision of the estimated efficiency scores and the resulting rankings of the firms studied. Horrace and Schmidt (2000) introduced two multiple comparison techniques (MCC and MCB) based on sampling theory statistics to this literature. In this paper, we add a Bayesian approach to the toolkit for measuring the precision of efficiency estimates and the ability to such estimates to accurately differentiate between the units being ranked.

After presenting the details of how to implement the Bayesian Multiple Comparison (BMC) approach, we presented an application to a panel of 43 U.S. electric utilities. Bayesian estimation of a distance function yields a set of technical efficiency estimates consistent with economic theory that provide an empirical ranking of the 43 firms. Application of the BMC approach then allows us to analyze which firms can truly be differentiated from which others at any desired level of probability. That is, we can make statements such as "there is an 99.6% probability that firm 31 is more efficient than all the firms in group 4" and "there is a 54.7% probability that all the firms in group 1 are more efficient than all the firms in group 3." The MCB approach of Horrace and Schmidt was also applied to the same technical efficiency estimates and provided some contrasting results in terms of the ability to differentiate firms on the basis of their TE scores.

We believe that the Bayesian results provide more flexibility in terms of multiple comparisons that are possible, particularly for those researchers who are not statistical experts. Using the procedure outlined in this paper, it is straightforward to compute the probability of any firm or group of firms being more efficient than any other firm or group of firms. This probability provides an exact measure of the ability to rank the groups/firms

according to their technical efficiency estimates (and the appropriate confidence in those rankings). While the MCB and MCC methods do allow for such flexibility, it is a much more complex matter to generalize the approach to compute such comparisons. Also, the sampling theoretic-based MCC and MCB approaches do not yield finite sample probability values to measure the differentiation between the TE scores of the firms. Instead, the method provides the normal (for sampling theory statistics) all or nothing test results where firms are either differentiated from the best (or index firm) or are not.

The greater information content and flexibility of the Bayesian approach are significant advantages in providing statistical information about the precision of efficiency rankings. Further, the method is more straightforward from a statistical viewpoint, requiring nothing more complicated than a basic ability to generate random numbers from known statistical distributions, a function available in nearly all of statistical and econometrics software packages on the market today.

The application presented here used a distance function framework with some attendant complications due to the presence of a bad output and endogeneity necessitating the use of an instrumental variables approach. However, applications of the Bayesian approach presented can be easily implemented for technical efficiency estimates from a stochastic frontier model which could be estimated in a simpler manner. Regardless of the approach, once the posterior distributions of the technical efficiency estimates have been derived (or numerically approximated), the Bayesian Multiple Comparison (BMC) approach presented here can be easily performed at little additional cost in terms of programming time and effort. In contrast to the simplicity of the approach, the information generated by BMC approach is quite rich. It yields considerable useful information for policy and decision makers who wish to know the accuracy and differentiability of estimated rankings.

Appendix: Pseudo Code for Step 1 of the Gibbs Sampler

a. First, generate a Wishart random variable $A^{(i)} = [\Sigma^{(i)}]^{-1} \sim W(S^{-1}, \nu)$.

   i. Let $LL' = S^{-1}$, where L is a lower triangular matrix from the Cholesky decomposition of S^{-1}. We obtain S^{-1} as the inverse of the estimated variance-covariance matrix of the error terms.

   ii. Assume that $Q \sim W(I, \nu)$. Then $LQL' \sim W(LL', \nu) = W(S^{-1}, \nu)$

   iii. Now from Anderson (1984) $UU' = Q \sim W(I, \nu)$, where $U$is lower triangular, all u_{ij} are independent, $u_{ij} \sim N(0, 1), i > j$, and $u_{ii}^2 \sim \chi^2(\nu - i + 1)$ random variable, which implies that $LUU'L' \sim W(S^{-1}, \nu)$. Thus, we draw u_{ij} values and form $Q^{(i)} = U^{(i)}U^{(i)'}$.

   iv. Now use L and $Q^{(i)}$ compute $LQ^{(i)}L'/\nu = A^{(i)}$.

b. Compute $\Sigma^{(i)} = [A^{(i)}]^{-1}$, which is an inverse-Wishart random variable.

References

Anderson, T.W., *An Introduction to Multivariate Statistical Analysis*, 2nd ed., New York: John Wiley (1984).

Atkinson, S. and C. Cornwell, Measuring and Decomposing Productivity Change: Stochastic Distance Function Estimation vs. DEA, Working Paper, University of Georgia (2000).

Atkinson, S. E., and J. H. Dorfman, Crediting Electric Utilities for Reducing Air Pollution: Bayesian Measurement of Productivity and Efficiency, Working Paper, University of Georgia (2001).

Baltagi, B. H, and J. M. Griffin, A General Index of Technical Change, *Journal of Political Economy* 96, 20-41 (1988).

Barbera, A. and V. McConnell, The Impact of Environmental Regulations on Industry Productivity: Direct and Indirect Effects, *Journal of Environmental Economics and Management* 18, 50-65 (1990).

Callan, S. J., 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 (1991).

Casella, G., and E. I. George, Explaining the Gibbs Sampler, *The American Statistician* 46, 167-174 (1992).

Caves, D. W., L. W. Christensen, and W. E. Diewert, The Economic Theory of Index Numbers and the Measurement of Input, Output, and Productivity, *Econometrica* 50, 1393–1414 (1982).

Chambers, R. G., Y. Chung, and R. Färe, Benefit and Distance Functions, *Journal of Economic Theory*, 70, 407–419 (1996a).

Chambers, R. G., Y. Chung, and R. Färe, Profit, Directional Distance Functions, and Nerlovian Efficiency, Working Paper, Department of Economics, Southern Illinois University (1996b).

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

Chung, Y. and R Färe, Productivity and Undesirable Outputs: A Directional Distance Function Approach, Working Paper, Department of Economics, Southern Illinois University (1997).

Conrad, K. and D. Wastl, The Impact of Environmental Regulation on Productivity in German Industries, *Empirical Economics* 20, 615–633 (1995).

Cornwell, C., P. Schmidt, and R.C. Sickles, Production Frontiers with Time Series Variation in Efficiency Levels, *Journal of Econometrics* 46, 185–200 (1990).

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

Färe, R., S. Grosskopf, C. Lovell, and C. Pasurka, Multilateral Productivity Comparisons When Some Outputs are Undesirable: A Nonparametric Approach, *Review of Economics and Statistics*, 71, 90–98 (1989).

Färe, R. and D. Primont, *Multi-Output Production and Duality: Theory and Applications,* Boston: Kluwer Academic Publishers (1995).

Fernandez, C., G. Koop, and M. Steel, Multiple-Output Production with Undesirable Outputs, working paper, Department of Economics, University of Edinburgh, 2000.

Geweke, J., Using Simulation Methods for Bayesian Econometric Models: Inference, Development and Communication (with discussion and rejoinder), *Econometrics Reviews* 18, 1-126 (1999).

Gollop, F. and M. Roberts, Environmental Regulations and Productivity Growth: The Case of Fossil-fueled Electric Power Generation, *Journal of Political Economy* 91, 655-674 (1983).

Hahn, G. J., and R. W. Hendrickson, A Table of Percentage Points of the Distribution of the Largest Absolute Value of k Student t Variables and its Applications, *Biometrika* 58, 323-332 (1971).

Hansen, L. P. , Large Sample Properties of Generalized Method of Moments Estimation, *Econometrica* 50, 1029-1054 (1982).

Horrace, W. C. and P. Schmidt, Multiple Comparisons with the Best, with Economic Applications, *Journal of Applied Econometrics* 15, 1-26 (2000).

Koop, G., Cross-Sectoral Patterns of Efficiency and Technical Change in Manufacturing, *International Economic Review* 42, 73–104 (2001).

Koop, G., Carbon Dioxide Emissions and Economic Growth: A Structural Approach, *Journal of Applied Statistics* 25, 489–515 (1998).

Nelson, R. A., Regulation, Capital Vintage, and Technical Change in the Electric Utility Industry, *Review of Economics and Statistics* 66, 59–69 (1984).

Pittman, R., Multilateral Productivity Comparisons with Undesirable Outputs, *The Economic Journal* 93, 883–891 (1983).

Reinhard, S. and Thijssen, G., Resource Use Efficiency of Dutch Dairy Farms: A Parametric Distance Function Approach, working paper, Wageningen Agricultural University (1999).

Tanner, M. A., *Tools for Statistical Inference* New York: Springer, (1996).

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

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

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

Yaisawarng, S. and J. Klein, The Effects of Sulfur Dioxide Controls on Productivity Change in the U.S. Electric Power Industry, *Review of Economics and Statistics* 76, 447–460 (1994).

Zellner, A., The Finite Sample Properties of Simultaneous Equations' Estimates and Estimators Bayesian and non-Bayesian Approaches, *Journal of Econometrics* 83, 185–212 (1998).

Zellner, A. and J. Tobias, Further Results on Bayesian Method of Moments Analysis of the Multiple Regression Model, *International Economic Review* 42, 121–140 (2001).

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	CEC 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: Time-Persistent Technical Efficiency Score by Firm

Utility	Tech. Eff. Score	Gibbs Efficiency Grouping	MCB Eff. Set
1	0.279544	4	
11	0.309533	4	
10	0.335788	4	
9	0.357404	4	
13	0.410859	3	*
16	0.421139	3	*
2	0.437480	3	
12	0.439326	3	
27	0.441344	3	
17	0.445462	3	*
3	0.452934	3	
22	0.484653	3	
34	0.487115	3	*
35	0.492744	3	*
41	0.496442	3	*
8	0.497090	3	*
33	0.501968	3	
6	0.511428	3	*
32	0.511551	3	*
20	0.513567	3	*
4	0.533391	3	
18	0.536600	2	*
42	0.536890	3	*
39	0.537757	2	*
21	0.554688	2	
43	0.558791	2	*
30	0.569756	2	
14	0.584080	2	*
36	0.590789	2	*
5	0.600555	2	
7	0.602033	2	*
40	0.610420	2	*
25	0.611875	2	*
37	0.614052	2	*
38	0.654482	2	*
23	0.665528	2	*
28	0.685584	1	*
15	0.713970	1	*
26	0.768943	1	*
24	0.770113	1	*
29	0.826387	1	*
19	0.851807	1	*
31	0.911533	1	*
Wtd. Avg.	0.551567		

Table 3: Bayesian Multiple Comparison Results

Group A	Group B	$\text{Prob}(\text{TE}_{\text{Group A}} \geq \text{TE}_{\text{Group B}})$
G3	G4	0.383
G2	G4	0.915
G1	G4	0.980
G2	G3	0.002
G1	G3	0.547
G1	G2	0.069
rest of G4	F1	0.641
F31	rest of G1	0.403
G3	F1	0.933
G2	F1	0.990
G1	F1	0.994
F31	G4	0.996
F31	G3	0.917
F31	G2	0.834

Figure 1: Technical Efficiency Rankings