



# Multiple Comparisons with the Best: Bayesian Precision Measures of Efficiency Rankings

SCOTT E. ATKINSON

atknsn@terry.uga.edu

*Department of Economics, University of Georgia, Athens, GA 30602, U.S.A.*

JEFFREY H. DORFMAN

jdorfman@agecon.uga.edu

*Department of Agricultural and Applied Economics, University of Georgia, Athens, GA 30602, U.S.A.*

## *Abstract*

A large literature measures the allocative and technical efficiency of a set of firms using econometric techniques to estimate stochastic production frontiers or distance functions. Typically, researchers compute only the precision of individual efficiency rankings. Recently, Horrace and Schmidt (*Journal of Applied Economics* 15, 1–26, 2000) have applied sampling theoretic statistical techniques known as multiple comparisons with a control (MCC) and multiple comparisons with the best (MCB) to make statistical comparisons of efficiency rankings. As an alternative, this paper offers a Bayesian multiple comparison procedure that we argue is simpler to implement, gives the researcher increased flexibility over the type of comparison, and provides greater, and more intuitive, information content. For these methods and a parametric bootstrap technique, we carry out multiple comparisons of technical efficiency rankings for a set of U.S. electric generating firms, estimated using a distance function framework. We find that the Bayesian method provides substantially more precise inferences than obtained using the MCB and MCC methods.

**JEL Classification:** C11, C32, D24

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

## **1. Introduction**

Situations abound in which economists, decision makers, and other interested parties desire a ranking of some set of performance measures. 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). Firms are ranked relative to intra-industry competitors on the basis of technical efficiency. In all these situations, in addition to

the ranking, it would be beneficial to provide information on the precision of the rankings and whether one firm's rank differs significantly from the ranks of other firms. That is, 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 be distinguishable due to a lack of statistical precision. In such a case, the ranking would be best suppressed.

A substantial 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 non-stochastic and thus the precision of these rankings can only be evaluated using bootstrapping methods after imposing a stochastic structure on the problem (see Atkinson and Wilson, 1995). A parallel literature uses econometric techniques, such as estimation of stochastic production frontiers or distance functions, thus providing at least the possibility of directly computing significant differences among efficiency rankings. Recently, Horrace and Schmidt (2000) have applied sampling theoretic statistical techniques known as multiple comparisons with a control (MCC) and multiple comparisons with the best (MCB) to the issue of comparing 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 indistinguishable from 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.<sup>1</sup> The Bayesian method easily allows comparisons between single firms, a firm versus a group, or one group versus another 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% posterior probability" are possible. We also consider a parametric bootstrap technique, which is computationally much simpler than our Bayesian approach. We compare technical efficiency rankings of a set of U.S. electric generating firms derived within a distance function framework. We find that the Bayesian inferences are considerably more precise than those obtained using the MCB and MCC procedures.

The remainder of this paper is organized as follows. In Section 2, we review the MCB and MCC approaches pioneered by Horrace and Schmidt to make multiple comparisons of efficiency rankings. In Section 3, we introduce a Bayesian alternative, discussing differences and potential advantages of this methodology. In Section 4 we discuss the model, the data, and the estimation procedure used to generate technical efficiency rankings. Section 5 compares results from three alternative methods. Conclusions follow in Section 6.

## 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 arguably is 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. This is assumedly due to the stochastic and imprecise nature of the estimated rankings. In many real world applications, 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  $\theta_i, i = 1, \dots, N$ . The MCC method computes a joint confidence interval of a desired probability level for all the differences between individual firm efficiencies and the index firm  $j$ , that is, for the vector  $[\theta_j - \theta_1, \theta_j - \theta_2, \dots, \theta_j - \theta_N]$ . When the efficiency estimates are correlated, this joint confidence interval is given just after equation (14) in Horrace and Schmidt (2000):

$$P\left(L_i^j \leq \theta_j - \theta_i \leq U_i^j \quad \forall i \neq j\right) = 1 - \alpha, \quad (2.1)$$

where  $L_i^j = \hat{\theta}_j - \hat{\theta}_i - h_{ji}$ ,  $U_i^j = \hat{\theta}_j - \hat{\theta}_i + h_{ji}$ , and  $h_{ji} = d_j^* \hat{\sigma}_{ji}$ , where  $d_j^*$  is the critical value for the joint two-sided confidence interval which has been adjusted to account for the multiple comparisons and  $\hat{\sigma}_{ji}$  is the standard deviation of  $\hat{\theta}_j - \hat{\theta}_i$ . (for details see Horrace and Schmidt, 2000, equation (12)). Tables of critical values for MCC can be found in Hahn and Hendrickson (1971), *inter alia*.

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 efficiency measure for the fixed index firm  $\theta_j$  with the efficiency measure of an unknown best index firm,  $\theta_{(N)}$ , in Horrace and Schmidt's notation. This unknown index firm must be a member of the "efficient set"  $S$  which is defined for probability level  $(1 - \alpha)$  as

$$S(1 - \alpha) = \{i : \theta_i \geq \theta_j - h_{ij} \quad \forall j \neq i\}, \quad (2.2)$$

with  $h_{ij}$  as defined in Section 2.1.

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. Intuitively, this is the set of firms whose estimated efficiency measures are not statistically significantly below the measure of the uncertain most efficient firm. Given the set  $S$  which contains the firms which could possibly be the best firm, one can construct a joint confidence interval for all firms' efficiency measures relative to this unknown best firm's efficiency measure. Such a confidence interval is defined by a set of limits  $\{(L_i, U_i) \quad \forall i\}$  such that

$$P[L_i \leq \theta_{(N)} - \theta_i \leq U_i \quad \forall i : (N) \in S(1 - \alpha)] \geq 1 - \alpha. \quad (2.3)$$

This joint confidence interval is exactly analogous to the MCC one except for the replacement of the fixed index firm with the uncertain best firm within the membership of set  $S$ . For more on the computation of the limits for both the MCC and MCB confidence intervals, the reader is encouraged to see Horrace and Schmidt (2000).

### 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 methodology provides exact (posterior) probability levels for each comparison statement to be evaluated (recall from (2.3) that the frequentist approach provides an MCB confidence interval with a  $\geq$  not an  $=$  sign). Further, the Bayesian approach allows for a variety of more general and flexible probability statements to be constructed and evaluated, allowing virtually any statistic of interest to be estimated.

The Bayesian approach is fairly simple to implement given the set of parameter draws, generated in our application by a Gibbs sampling algorithm, which modern Bayesians use to compute posterior estimates of the unknown parameters (in the current situation: the technical efficiency scores). As will be detailed precisely in Section 3.1, 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.

Given a set of random draws from the posterior density function of a vector of parameters,  $\gamma$ , one can estimate the posterior mean of a function of interest, say  $g(\gamma)$ , 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 draw from the posterior distribution 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, and standard deviations, they can be compared across firms. To estimate the probability that firm  $i$  is more efficient than firm  $j$ , we compute the proportion of draws for which firm  $i$ 's TE score is greater. Formally, for a set of  $B$  draws on the TE scores,

$$P(TE_i > TE_j) = B^{-1} \sum_{b=1}^B H \left( TE_i^{(b)} > TE_j^{(b)} \right), \quad (3.1)$$

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  $TE_i^{(b)} > \max\{TE_j^{(b)}, j \in J\}$ .

### 3.1. The Bayesian Efficient Set

The above leads naturally to the Bayesian computation of the efficient set  $S$ . First, compute the probability that each firm is the "best" firm,

$$P(TE_i > \max\{TE_j \forall j \neq i\}) = B^{-1} \sum_{b=1}^B H \left( TE_i^{(b)} > \max\{TE_j \forall j \neq i\} \right), \quad (3.2)$$

where  $H$  is again the logical operator equal to one when the argument is true and zero otherwise.

To turn the individual firm probabilities of being the "best" firm into an efficient set  $S$ , simply order the firms from lowest probability to highest and include firms in the set  $S$  until the cumulative probability reaches the desired level. Define the ordered set of the probabilities from equation (3.2) as  $\{\omega_{(i)}, \forall (i) = (1), \dots, (N)\}$ .

Mathematically, the Bayesian definition of the efficient set of potential best firms becomes

$$S(1-\alpha) = \left\{ (i) : (i) \geq (j), \sum_{(i)=(j)}^{(N)} \omega_{(i)} \geq 1-\alpha, \text{ and } \sum_{(i)=(j+1)}^{(N)} \omega_{(i)} < 1-\alpha \right\}. \quad (3.3)$$

For a preselected probability level  $(1-\alpha)$ , the set  $S$  will have slightly more than the desired probability of containing the best firm given the discrete nature of including firms. In the Bayesian approach, the exact probability content of the set  $S$  is explicit when the desired probability level is achieved through cumulative summation. In the frequentist case, to obtain the analogous probability content, one would have to incrementally raise the probability level of the efficient set until a new firm enters this set. This implicitly defines the probability content of the original efficient set as equal to this new level minus an arbitrarily small  $\delta > 0$ .

### 3.2. Bayesian MCC

To create a Bayesian equivalent of the frequentist MCC joint confidence interval one must make a few choices of how to trade off the various interval limits, since (in both Bayesian and frequentist cases) the joint confidence interval is not unique. For example, should intervals be symmetric or the shortest possible; how do you measure the length of a joint confidence interval that is often highly multidimensional? These issues will be made clear as they are developed and used below.<sup>2</sup>

To begin the construction of a Bayesian MCC interval for a fixed index firm  $j$ , first construct vectors of differences for each efficiency measure from the index measure for all the draws that were generated, then sort each vector from smallest to largest. A 100% MCC joint confidence interval would take the ends of these ordered vectors as the necessary limits  $(L_i^j, U_i^j)$ .

To reduce the probability content of the MCC interval, we chose to discard symmetry in favor of finding the shortest possible joint interval, with shortest defined as the minimum sum of  $U_i^j - L_i^j$  across the firms  $i = 1, \dots, N$ . Computationally, that simply involves finding the firm  $i$  and limit (either upper or lower) such that the length of the joint confidence interval is reduced by the maximum possible amount. Candidate upper and lower limit changes are associated with firms  $m$  and  $n$ , respectively,

$$\tilde{L}_m^j = \operatorname{argmax} \left\{ \left( \text{TE}_j^{(2)} - \text{TE}_m^{(2)} \right) - \left( \text{TE}_j^{(1)} - \text{TE}_m^{(1)} \right) \forall m \right\}, \quad (3.4)$$

$$\tilde{U}_n^j = \operatorname{argmax} \left\{ \left( \text{TE}_j^{(B)} - \text{TE}_n^{(B)} \right) - \left( \text{TE}_j^{(B-1)} - \text{TE}_n^{(B-1)} \right) \forall n \right\}. \quad (3.5)$$

Once these are found, simple comparison determines which change in limits produces the biggest drop in the length of the joint confidence interval. Then, the appropriate upper or lower limit, either  $L_m^j$  or  $U_n^j$ , is moved to exclude that one

draw; that is, if the lower limit is the best choice, the new limit would now be

$$L_m^j = \left( \text{TE}_j^{(2)} - \text{TE}_m^{(2)} \right).$$

This process continues incrementally until  $\alpha B$  draws have been excluded by moving the limits inward. At the end of the process, one will have a set of  $(L_i^j, U_i^j)$  which define the shortest joint confidence interval containing  $(1 - \alpha)$  posterior probability of the differences in TE measures as measured from the index firm  $j$ . Such an interval is the Bayesian equivalent of an MCC interval.

While again noting that this MCC interval is not unique (like the frequentist one), it has some interesting and appealing features. First, it is the shortest possible joint confidence interval. Second, and related to the first point, the probabilities of each firm's efficiency difference being in the interval are not equal; that is, we do not simply have a set of  $(1 - \alpha)^{1/(N-1)}$  marginal confidence intervals. Excluding equal amounts of tail probability from each marginal confidence interval would define a different valid joint confidence interval. The standard, frequentist MCB interval excludes equal (and symmetric) tail probabilities; such a balanced interval will be necessarily equal to or longer than the shortest possible interval constructed without the constraint that the tail probabilities be balanced and equal. Within the potential set of valid joint confidence intervals, we argue that the shortest interval is the most likely to be preferred by policy makers. Generally, they desire the most precise inference possible and hence would prefer the shortest possible (joint) confidence interval.

### 3.3. Bayesian MCB

The creation of a Bayesian MCB joint confidence interval is straightforward given the above approach to the construction of an MCC interval. The fixed index firm needs to be replaced with the uncertain best firm from within the efficient set  $S$ . This is accomplished in the step of creating the ordered vectors of efficiency differences for each firm. Instead of taking differences between each firm's efficiency measure and the fixed index firm, for each draw we first establish the best firm from within  $S$  by finding the maximum value of  $\text{TE}_j^{(b)}$  across the firms  $j \in S$  for each draw  $(b) = 1, \dots, B$ . Then differences are taken between each firm's technical efficiency measure for each draw and the corresponding value of the best firm for that draw. From that point on, the construction of the MCB interval is identical to the method used to find the MCC interval.

### 3.4. Other Bayesian Possibilities

Horrace and Schmidt (2000) do not perform comparisons between groups or of a single firm versus another single firm or subgroup. While the frequentist MCC and MCB approaches can be extended to accomplish such tasks, they are undoubtedly easier to accomplish within a numerical Bayesian framework. Bayesian numerical methods make possible computation of posterior probabilities or confidence intervals (posterior density regions) that include the following possibilities (and more).

Following Fernández et al. (2002), one can compute the probability a firm belongs in a particular percentile group (the top quartile, the second decile, etc.). One can compute the probability a (set of) firm(s) is more efficient than another (set of) firm(s). One can produce a posterior density region (PDR) for the difference in efficiency between two firms, between one firm and the uncertain best of a group of other firms, or between the uncertain best firms from two distinct groups. One can construct a PDR similar to the MCC and MCB cases above, but with a single set of limits  $(L, U)$ . Such “cubical” intervals (actually equilateral hyperplanes) would be much easier for policy makers to absorb and these intervals make sense in applications where the majority of the firm-specific limits are not too dissimilar.

While frequentist approaches could be constructed for all intervals mentioned above, the statistical foundation of the 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 (requiring a new computer simulation of the critical values for each interval). Thus, while there is no theoretical barrier to stop MCC and MCB approaches from performing the same sorts of comparisons as our Bayesian approach, the task is more daunting than the Bayesian approach until more work is done to develop user-friendly software. Further, the frequentist approach cannot produce posterior probability statements due to the fundamental difference between the two branches of statistics. Frequentist  $p$ -values associated with hypothesis tests are not convertible into probability statements in favor of the null or alternative hypotheses.

#### 4. Empirical Application

We apply our methodology to a panel of U.S. electric utilities observed at five-year intervals from 1980 to 1995. There are two outputs: the quantity of electric power generated (a good output) and the quantity of SO<sub>2</sub> emissions (a bad output which locally has a direct negative effect on health and welfare and regionally can lead to acid rain). Capital, labor, and energy inputs are applied to produce the good and bad outputs. This application is particularly relevant since allowable SO<sub>2</sub> emissions from electric utilities have been reduced dramatically over the last decade and since electricity is sporadically in short supply in the State of California, where State Implementation Plans are very strict. Title IV of the 1990 Clean Air Act Amendments required reductions in emissions of SO<sub>2</sub> from U.S. coal-burning electric utilities from about 19 million tonnes in 1980 to 8.95 million tonnes by the year 2000. Utilities should be credited for this increased reduction of SO<sub>2</sub> emissions over time.

##### 4.1. The Model

Let  $\mathbf{x}$  be a vector of inputs denoted by  $\mathbf{x} = (x_1, \dots, x_N) \in R_+^N$ ,  $\mathbf{y}$  be a vector of good outputs denoted by  $\mathbf{y} = (y_1, \dots, y_M) \in R_+^M$ , and  $\mathbf{b}$  be a vector of bad outputs

denoted by  $\mathbf{b} = (b_1, \dots, b_Z) \in R_+^Z$ . Using an input distance function, symmetric treatment of bad outputs and good outputs can be specified using an input distance function, where such treatment is emphasized with square brackets

$$D_i([\mathbf{y}, \mathbf{b}], \mathbf{x}, t) = \sup_{\lambda} \{ \lambda : ([\mathbf{y}, \mathbf{b}], \mathbf{x}/\lambda) \in S(\mathbf{x}, \mathbf{b}, \mathbf{y}, t) \}. \quad (4.1)$$

The goods and bads are held constant while inputs are proportionally scaled downward to their minimum required level.<sup>3</sup> Since the input distance function in (4.1) is dual to the cost function, we can write, following Atkinson and Primont (2002),

$$C_i([\mathbf{y}, \mathbf{b}], \mathbf{p}, t) = \min_{\mathbf{x}} \{ \mathbf{p}\mathbf{x} : D_i([\mathbf{y}, \mathbf{b}], \mathbf{x}, t) \geq 1 \}, \quad (4.2)$$

where  $\mathbf{p} = (\mathbf{p}_1, \dots, \mathbf{p}_N) \in R_+^N$  is a vector of input prices and  $C([\mathbf{y}, \mathbf{b}], \mathbf{p}, t)$  is a unit cost function if costs are minimized. 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

$$\mathbf{p} = C([\mathbf{y}, \mathbf{b}], \mathbf{p}, t) \nabla_{\mathbf{x}} D_i([\mathbf{y}, \mathbf{b}], \mathbf{x}, t). \quad (4.3)$$

The appropriate monotonicity condition for the bad in the context of the input distance function can be derived following Atkinson and Dorfman (2004) to obtain

$$\frac{\partial D_i}{\partial b} \geq 0. \quad (4.4)$$

As a flexible approximation to the true distance function in (4.1), we adopt the translog functional form. Arbitrarily letting the distance equal unity for all data points,  $\ln(1) = 0$  and the empirical model for firm  $f = 1, \dots, F$  in period  $t = 1, \dots, T$  has the form

$$\begin{aligned} 0 = & \gamma_0 + \sum_m \gamma_m \ln y_{mft} + \sum_z \gamma_z \ln b_{zft} + \sum_n \gamma_n \ln x_{nft} + \gamma_{t1}t + (1/2)\gamma_{t2}t^2 \\ & + (1/2) \sum_m \sum_{m'} \gamma_{mm'} \ln y_{mft} \ln y_{m'ft} + (1/2) \sum_z \sum_{z'} \gamma_{zz'} \ln b_{zft} \ln b_{z'ft} \\ & + (1/2) \sum_n \sum_{n'} \gamma_{nn'} \ln x_{nft} \ln x_{n'ft} + \sum_m \sum_n \gamma_{mn} \ln y_{mft} \ln x_{nft} \\ & + \sum_z \sum_n \gamma_{zn} \ln b_{zft} \ln x_{nft} + \sum_z \sum_m \gamma_{zm} \ln b_{zft} \ln y_{mft} \\ & + \sum_m \gamma_{mt} \ln y_{mft}t + \sum_z \gamma_{zt} \ln b_{zft}t + \sum_n \gamma_{nt} \ln x_{nft}t + \ln h(\epsilon_{ft}), \end{aligned} \quad (4.5)$$

where

$$h(\epsilon_{ft}) = \exp(v_{ft} - u_{ft}), \quad (4.6)$$

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.<sup>4</sup>

While the  $u_{ft}$  can be treated as fixed or random, the choice between the two entails a tradeoff. With the fixed effects approach, identification can be difficult, since the number of parameters increases with the number of firms,  $F$ . To identify the  $u_{ft}$  for each  $f$  and  $t$ , we use the model for time-varying inefficiency proposed by Cornwell et al. (1990):

$$u_{ft} = \sum_{q=0}^Q \beta_{fq} d_f t^q, \quad f = 1, \dots, F, \quad (4.7)$$

where  $t$  is a trend and  $d_f$  is a dummy variable equal to one for firm  $f$  and zero for the other firms.<sup>5</sup> With a fixed effects approach, the  $\beta_{fq}$  are firm-specific parameters to be estimated. This avoids the distributional and exogeneity assumptions that would otherwise be required in a random effects setup. The estimated distance equation is obtained by substituting (4.7) into (4.6), which in turn is substituted into (4.5), so that the  $\beta_{fq}$  are fit directly with the other parameters. Our estimated system consists of the distance equation plus a price equation as in (4.3) for each input, all of which are assumed to have symmetric error terms with zero means. We term this set of equations our input distance system.

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.

Because 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 (4.5) (i.e., those terms other than  $h(\epsilon_{ft})$ ), adding and subtracting  $\hat{u}_t$  yields

$$0 = \ln \hat{D}_i(\mathbf{y}, \mathbf{x}, t) + \hat{v}_{ft} - \hat{u}_{ft} + \hat{u}_t - \hat{u}_t = \ln \hat{D}_i^*(\mathbf{y}, \mathbf{x}, t) + \hat{v}_{ft} - \hat{u}_{ft}^*, \quad (4.8)$$

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

Given (4.7), we estimate firm  $f$ 's level of technical efficiency in period  $t$ ,  $TE_{ft}$ , as

$$TE_{ft} = \exp(-\hat{u}_{ft}^*), \quad (4.9)$$

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

Before estimation of our distance system, several sets of parametric restrictions are imposed on (4.5) and (4.3): symmetry, linear homogeneity in input quantities, and the identifying constraint that  $\beta_{1q} = 0$ ,  $\forall q$  in (4.7).

**4.2. Data**

Our dataset is an updated and refined version of the panel of utilities originally analyzed by Nelson (1984).<sup>6</sup> 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 (see Atkinson and Dorfman, 2004 for details).

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<sub>2</sub> emissions ( $b$ ) are measured in tonnes. We also employ a number of instruments in our Generalised Method of Moments (GMM) procedure; details are provided in Atkinson and Dorfman (2004).

**4.3. Bayesian Estimation Procedure**

We employ a limited information Bayesian system estimator for the estimation of our distance system, rather than a full-information Bayesian estimator which has the inherent difficulties both in implementation and in accurate specification of the full likelihood including issues of endogeneity. We follow and generalize Kim (2002) and Zellner and Tobias (2001) to provide a limited-information Bayesian estimation algorithm. Kim shows that maximizing entropy subject to a restriction on the GMM criterion function yields an optimal limited-information likelihood function (LILF). We generalize Kim’s approach by treating the covariance of the errors as well as unknown parameters of our distance system as random variables and constructing a joint LILF for both.<sup>7</sup> The least informative LILF,  $f$ , can be found by solving the following optimization problem

$$\operatorname{argmax}_{f \in \mathcal{F}} - \int f(\gamma, \Omega | D) \ln f(\gamma, \Omega, D) d\gamma d\Omega, \tag{4.10}$$

where  $\gamma$  is a vector of regression model parameters,  $\Omega$  is the covariance matrix of the regression model’s stochastic error terms, and  $D$  represents the data. The solution is given by

$$\hat{f}(\gamma, \Omega | D) = c_0 |\Omega|^{-c_1} \exp \left[ -c_2 h(\gamma)' S^{-1} h(\gamma) - c_3 \operatorname{tr}(\Xi \Omega^{-1}) \right], \tag{4.11}$$

which is the product of a distribution from the exponential family for  $\gamma$  and an inverted Wishart with respect to  $\Omega$  (Zellner and Tobias, 2001), where  $c_0$ ,  $c_1$ ,  $c_2$ , and  $c_3$  are constants,  $S = E[h(\gamma)h(\gamma)']$  and  $\Xi$  is the sum of squared errors matrix (see Atkinson and Dorfman, 2004 for details). One can introduce a prior density  $p(\gamma, \Omega)$  exactly as in any Bayesian inference problem, apply Bayes Theorem using the LILF in place of a standard likelihood function, and derive a

limited-information posterior distribution

$$f(\gamma, \Omega|D) = p(\gamma, \Omega) \hat{f}(\gamma, \Omega|D) c^{-1}, \quad (4.12)$$

where  $c$  is the normalizing constant.

The prior distribution used in our application 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. This prior distribution can be written as

$$p(\gamma, \Omega) \propto \text{MVN}(g_o, H_o) |\Omega^{-1}|^{-5/2} I(\gamma, \mathcal{R}), \quad (4.13)$$

where MVN is the multivariate normal distribution,  $g_o$  is the vector of prior means on the parameters in  $\gamma$ ,  $H_o$  is the prior variance–covariance matrix on the same parameters,  $I(\gamma, \mathcal{R})$  represents the indicator function that equals one when the restrictions are satisfied and zero otherwise, and the exponent 5/2 is equal to  $(m+1)/2$  where  $m=4$ , the number of equations in our system.

The vector  $g_o$  is set to zero and  $H_o$  is set to 100. This is a very diffuse prior, having virtually no effect on the posterior means, but ensures that the prior is proper in any dimensions that are not restricted to a finite subspace by the indicator function part of the prior. The prior for  $\Omega$  (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 output, and the bad output. Because monotonicity is a point-specific condition for our functional form and to allow for 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.<sup>8</sup>

Given a limited-information likelihood in the form of (4.11) and the prior in (4.13), we can obtain the posterior distribution. Given constants that are embedded in the restrictions on the GMM criterion function, the limited-information posterior takes the form of

$$p(\gamma, \Omega|D) = c_o |\Omega|^{-(n-k+m+1)/2} \exp \left[ -\frac{1}{2} (\gamma - \gamma_p)' \Psi_p^{-1} (\gamma - \gamma_p) - \frac{1}{2} \text{tr} (\Xi \Omega^{-1}) \right] I(\gamma, \mathcal{R}), \quad (4.14)$$

where

$$\gamma_p = \Psi_p (H_o^{-1} g_o + \Psi_m^{-1} \gamma_m), \quad (4.15)$$

$$\Psi_p = (H_o^{-1} + \Psi_m^{-1})^{-1} \quad (4.16)$$

and  $\Psi_m$  is the standard GMM covariance matrix of  $\gamma_m$ , which is the standard GMM estimator of  $\gamma$  given the set of identifying restrictions specified. The

limited-information posterior distribution in (4.14) is a truncated version of the standard multivariate normal–inverted Wishart distribution common in Bayesian econometrics (for details see Atkinson and Dorfman, 2004).

## 5. Results

We analyzed the full joint limited-information posterior distribution using Gibbs sampling. The Gibbs sampler begins with initial values for  $\gamma$  and  $\Omega$  (we use GMM estimates for this purpose). Then a series of draws are made from the specified conditional posterior distributions for  $\Omega$  and  $\gamma$ , each conditioned on the most recent random draw of the other. Details are provided in Appendix A. Such a process converges to a random sample from the full joint posterior distribution as in Chib (1995).<sup>9</sup> After the first 500 draws were discarded to remove dependence on the initial conditions, we then drew 10,000 more parameter vectors for computation of the posterior distribution. Computed posterior standard deviations proved this number of draws to be sufficient. As a test of 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 the different means were found to be statistically equivalent, we conclude that our Gibbs sampler converged. Estimated TEs for all 43 firms in the sample are displayed in Table 1.

Using the estimates obtained above, we produce multiple comparisons by three different approaches. First, the Gibbs draws were used to compute the Bayesian efficient set and joint MCC and MCB intervals as described in Section 3. Second, the Bayesian posterior means and variances of the TE measures were used to compute Horrace and Schmidt's efficient set, MCC, and MCB confidence intervals. An alternative would have been to use classical GMM estimates to compute the Horrace–Schmidt intervals. However, these estimates do not satisfy the monotonicity restrictions, and their use would have introduced possible bias in the comparison of methods. Finally, for the parametric bootstrap, GMM estimates and a multivariate normal distribution truncated by the economic restrictions discussed in section 4 were used to produce an empirical distribution of TE estimates similar to our numerical Bayesian approach. Specifically, after initial classical GMM estimation of our distance system, we draw repeatedly from a multivariate normal distribution with mean equal to the estimated parameters and covariance matrix equal to the estimated covariance matrix. We then discard any draws that fail to satisfy our previously discussed monotonicity standards (see Krinsky and Robb, 1986 for details). The resulting empirical distribution was then used to produce multiple comparisons in exactly the same manner as with the Bayesian posterior distribution.<sup>10</sup>

Use of a more traditional non-parametric bootstrap method proved unworkable. One would have to reformulate (4.5) as a reduced-form equation. In order to impose linear homogeneity in prices, the left-hand side of equation (4.5) would have to be redefined as  $\ln(1/x^*)$ , where  $x^*$  is an arbitrarily chosen exogenous input. All input quantities on the right-hand side of (4.5) would also have to be

Table 1. Technical efficiency scores and efficiency sets.

Firm	Tech. Eff. Score	Prob(best)	Eff. Sets
1	0.279544	0	
11	0.309533	0	
10	0.335788	0	
9	0.357404	0	
13	0.410859	0.0001	
16	0.421139	0	
2	0.437480	0	
12	0.439326	0	
27	0.441344	0	
17	0.445462	0	
3	0.452934	0	
22	0.484653	0.0028	F
34	0.487115	0.0002	F
35	0.492744	0.0005	F
41	0.496442	0.0010	F
8	0.497090	0	
33	0.501968	0.0003	F
6	0.511428	0.0003	F
32	0.511551	0.0004	F
20	0.513567	0.0002	F
4	0.533391	0.0055	F
18	0.536600	0.0005	F
42	0.536890	0.0032	F
39	0.537757	0.0017	F
21	0.554688	0	F
43	0.558791	0.0039	F
30	0.569756	0.0003	F
14	0.584080	0.0010	F
36	0.590789	0.0047	F
5	0.600555	0.0002	F
7	0.602033	0.0001	F
40	0.610420	0.0057	F
25	0.611875	0.0005	F
37	0.614052	0.0014	F
38	0.654482	0.0018	F
23	0.665528	0.0114	F
28	0.685584	0.0091	F B
15	0.713970	0.0354	F B BT
26	0.768943	0.0229	F B
24	0.770113	0.1370	F B
29	0.826387	0.0930	F B BT
19	0.851807	0.2750	F B BT
31	0.911533	0.3799	F B BT
Wtd. Avg.	0.551567		

Note: F = in frequentist efficiency set, B = in Bayesian efficiency set, BT = in bootstrap efficiency set.

normalized by  $x^*$ . In general, the difficulty with this approach is that all inputs may be endogenous. Also, input prices on the left-hand side of the estimated price equations may also be endogenous. In either case, a reduced form cannot be obtained.

It should be noted that this paper presents an introduction and exposition on how a Bayesian approach can be employed to perform multiple comparisons among efficiency measures. This is not a report of a Monte Carlo experiment designed to neutrally test three competing approaches and measure their relative effectiveness and accuracy. Thus, we have no “true” efficiency measures against which to compare our empirical results. The Bayesian results, given their finite sample properties, their provision of the full posterior distribution of statistics of interest, and the almost diffuse nature of the prior used (other than the imposition of the restrictions from economic theory), will be taken as the baseline results. Results from the Horrace–Schmidt method and from the parametric bootstrap will be compared to the Bayesian results.

### ***5.1. Efficient Sets and the Probability of Being the Best***

In the Bayesian context, the construction of the efficient set is straightforward as discussed in Section 3. Taking 95% as the probability content desired of the efficient set, the empirical result is an efficient set of seven firms containing a posterior probability of including the best firm equal to 0.9523. The parametric bootstrap yields an efficient set constructed in a similar manner that contains only five firms and a stated probability content of 0.9540. Finally, the frequentist approach produces an efficient set of 31 firms with a probability level of 0.95 or greater. The firms included in the efficient sets and the Bayesian probabilities of each firm being the best are displayed in Table 1.

The most likely explanation for the vast difference in the size of the efficient sets between the frequentist approach and the Bayesian one is the different starting points. The Bayesian (and bootstrap) approach begins with no firms in the efficient set and adds firms until the probability content reaches the set level. The frequentist approach begins with all firms in the efficient set and then excludes those for which the null hypothesis of equal efficiency with the unknown best firm can be rejected. That is, any statistical uncertainty is resolved in favor of including firms. Thus, if the statistical test has low power, the efficient set will obviously contain additional firms that are not in the “true” efficient set. If the Bayesian results are taken as exact, the frequentist efficient set has a 0.9999 probability of containing the best firm. For comparison, the bootstrap results which have a 0.9523 probability of containing the best firm under its own estimated probabilities, contains a 0.8019 probability under the Bayesian results. Thus, relative to the Bayesian results, the bootstrap efficient set is too small and the frequentist is too large. Presumably, the bootstrap efficient set is too small because this method draws coefficient values dependent on a given covariance matrix, rather than a distribution of covariance matrices as with the Bayesian procedure.

Table 2. MCB Interval results.

Firm	Bayesian		Frequentist		Bootstrap	
	$L_i$	$U_i$	$L_i$	$U_i$	$L_i$	$U_i$
1	0.000	0.897	0.000	1.064	0.000	0.843
2	0.000	0.806	0.000	0.954	0.000	0.742
3	0.000	0.832	0.000	0.958	0.000	0.686
4	-0.155	0.772	0.000	0.916	0.000	0.724
5	0.000	0.633	0.000	0.726	0.000	0.550
6	0.000	0.770	0.000	0.934	0.000	0.693
7	0.000	0.697	0.000	0.769	0.000	0.584
8	0.000	0.767	0.000	0.898	0.000	0.698
9	0.000	0.837	0.000	1.002	0.000	0.781
10	0.000	0.860	0.000	1.016	0.000	0.817
11	0.000	0.895	0.000	1.045	0.000	0.834
12	0.000	0.790	0.000	0.936	0.000	0.733
13	0.000	0.840	0.000	0.996	0.000	0.771
14	0.000	0.692	0.000	0.768	-0.061	0.498
15	0.000	0.649	0.000	0.745	0.000	0.530
16	0.000	0.831	0.000	0.987	0.000	0.750
17	0.000	0.815	0.000	0.960	0.000	0.719
18	0.000	0.766	0.000	0.849	0.000	0.607
19	0.000	0.619	0.000	0.715	0.000	0.433
20	0.000	0.802	0.000	0.950	0.000	0.751
21	0.000	0.729	0.000	0.826	0.000	0.686
22	-0.003	0.835	0.000	0.991	0.000	0.799
23	-0.094	0.687	0.000	0.798	0.000	0.586
24	0.000	0.671	0.000	0.793	-0.092	0.616
25	0.000	0.691	0.000	0.766	0.000	0.564
26	0.000	0.567	0.000	0.569	-0.108	0.392
27	0.000	0.810	0.000	0.952	0.000	0.812
28	0.000	0.654	0.000	0.757	0.000	0.610
29	0.000	0.544	0.000	0.545	0.000	0.448
30	0.000	0.749	0.000	0.873	0.000	0.705
31	0.000	0.511	0.000	0.426	0.000	0.268
32	0.000	0.798	0.000	0.937	0.000	0.745
33	0.000	0.777	0.000	0.938	0.000	0.756
34	0.000	0.786	0.000	0.933	0.000	0.708
35	0.000	0.807	0.000	0.959	0.000	0.743
36	-0.028	0.751	0.000	0.898	-0.000	0.675
37	-0.012	0.702	0.000	0.836	0.000	0.646
38	-0.022	0.691	0.000	0.802	0.000	0.648
39	0.000	0.790	0.000	0.937	0.000	0.712
40	-0.072	0.728	0.000	0.844	0.000	0.590
41	0.000	0.828	0.000	0.978	0.000	0.775
42	-0.022	0.773	0.000	0.939	0.000	0.698
43	-0.051	0.792	0.000	0.947	0.000	0.725

### 5.2. Comparing MCB Results

The MCB intervals were computed using each of the three methods being compared. The resulting upper and lower limits are shown in Table 2. Due to the magnitude of this information and the difficulty of comparing 43-dimensional intervals across three methods, we have summarized the results in two ways.

First, the “length” of the MCB intervals was computed using two different vector norms. The one-norm is the sum of the differences between each firm’s  $U_i$  and  $L_i$ ; that is the sum of the 43 separate firm-specific interval lengths.<sup>11</sup> The two-norm (or Euclidean norm) is the square root of the sum of the squared lengths of each firm’s interval length. The one- and two-norms for each method are displayed in Table 3. These results clearly show that the bootstrap gives the shortest intervals and the frequentist yields the longest.

The second method of summarizing the results was to compute coverage rates. Again taking the Bayesian posterior distributions as the standard, we can compute the empirical coverage rate of the MCB intervals and compare that to the stated value of 0.95. We find that the frequentist MCB interval has an empirical coverage of 0.9919 and the bootstrap has an empirical coverage of 0.5673. The larger than desired coverage of the frequentist MCB interval is not surprising and has been alluded to in that literature. In fact, Horrace and Schmidt (2000) clearly state that the MCB is a set that has probability  $p \geq (1 - \alpha)$  not  $p = (1 - \alpha)$ . The undersized nature of the bootstrap results is surprising in its magnitude. The vector norm results suggested that the intervals were too small, but did not show the true magnitude of the problem. By producing results conditional on a particular covariance matrix, the bootstrap approach understates the true parameter uncertainty and produces an MCB interval that is much too narrow. In doing the Gibbs sampling for the Bayesian estimation, we found the variation in the covariance matrix to be extremely important. The correlation between the covariance matrix and the structural parameters (and assumedly, through that, the efficiency estimates) is quite high.

### 5.3. Comparing Bayesian and MCB Results from Other Studies

Kim and Schmidt (2000) is the only other study we are aware of that compares Bayesian, classical MCB, and bootstrap coverage intervals. They estimate stochastic frontier models for three different data sets and also find that the MCB method produces the widest coverage intervals. Whether the coverage interval of the MCB

Table 3. Lengths and coverage rates of MCB intervals.

Method	1-Norm	2-Norm	Coverage rate
Frequentist	37.43	5.77	99.19%
Bayesian	32.70	5.02	95.00%
Bootstrap	28.91	4.47	56.73%

method is substantially larger than that of the Bayesian method depends on which model (fixed or random) is examined. With all three data sets, classical MCB and Bayesian coverage intervals are very similar for the random-effects model, but the classical MCB intervals are often substantially larger than the Bayesian ones for the fixed-effects model. In this sense, the results of Kim and Schmidt (2000) are consistent with those presented in this paper, since we estimate a fixed-effects model.

Two additional factors could explain differences between coverage intervals of the Bayesian and classical MCB models for our fixed effects model and theirs. First, the size of the efficient set varied widely among methods in our empirical example; Kim and Schmidt (2000) do not report their efficient sets, so we are not sure if they had a similar disparity in the size of the efficient sets. Clearly, a larger efficient set (as we find using the classical method) will lead to larger MCB intervals. Second, our Bayesian intervals are shortest intervals, while the Kim and Schmidt (2000) intervals are symmetrical.

#### **5.4. Comparing our MCB and MCC Results**

The MCC intervals were computed for all firms as the index firm using all three of the methods being compared. To save space, these results are not reported in detail. Briefly summarized, the MCC intervals are very similar to the MCB intervals in terms of relative length and coverage rates. Interested readers can obtain full results from the authors.

#### **5.5. Other Results with a Bayesian Approach**

Among some of the other results that the Bayesian approach can produce is the probability of a firm exceeding a particular ranking (or, equivalently, a specific order statistic). Following Fernández et al. (2002), the Bayesian joint posterior distribution of the efficiency measures can be used to compute the probability that a specific firm is in the top quartile, is more efficient than the median firm, and other similar comparisons. A selection of these results are displayed in Table 4. Comparing each firm's point estimate of its rank ordering in technical efficiency with its probability of exceeding each of the quartiles shown provides insight into the relative precision of the different posterior estimates of technical efficiency. For example, firm 31 is the most efficient firm according to the posterior means, but does not have the highest probability of being in the upper quartile. The highest probability of being in the upper quartile belongs to firm 29 at 0.970, while that for firm 31 is 0.951.

The Bayesian approach also lends itself to more flexibility and to the design of any "customized" probability statements that are desired. For example, in some cases, it may be useful to have single limit joint confidence intervals; that is, statements of the form

$$P[(N) \in S(1 - \alpha) \text{ and } L \leq \theta_{(N)} - \theta_i \leq U] = 1 - \alpha, \quad (5.1)$$

*Table 4.* Probability of firm's efficiency ranking exceeding a given quartile.

Firm	Lower quartile	Median	Upper quartile
1	0.003	0.000	0.000
2	0.413	0.067	0.004
3	0.493	0.129	0.011
4	0.847	0.455	0.113
5	0.983	0.774	0.288
6	0.850	0.307	0.032
7	0.985	0.794	0.259
8	0.793	0.253	0.009
9	0.018	0.000	0.000
10	0.009	0.000	0.000
11	0.023	0.003	0.001
12	0.408	0.038	0.000
13	0.198	0.020	0.002
14	0.928	0.685	0.252
15	1.000	0.982	0.789
16	0.282	0.040	0.004
17	0.454	0.073	0.004
18	0.871	0.495	0.068
19	0.974	0.918	0.806
20	0.876	0.317	0.031
21	0.957	0.625	0.043
22	0.613	0.248	0.064
23	0.999	0.958	0.567
24	1.000	0.995	0.883
25	0.987	0.842	0.294
26	0.999	0.987	0.892
27	0.453	0.039	0.000
28	0.999	0.982	0.708
29	1.000	0.998	0.970
30	0.990	0.738	0.108
31	0.998	0.989	0.951
32	0.862	0.306	0.038
33	0.774	0.274	0.042
34	0.740	0.210	0.017
35	0.741	0.222	0.028
36	0.989	0.769	0.216
37	0.998	0.906	0.288
38	1.000	0.991	0.545
39	0.920	0.467	0.089
40	0.986	0.823	0.295
41	0.721	0.251	0.048
42	0.903	0.456	0.102
43	0.964	0.574	0.139

or

$$P[L \leq \theta_{(\text{best})} - \theta_i \leq U] = 1 - \alpha. \quad (5.2)$$

The above equations define MCB confidence intervals with only a single set of limits—the  $U$  and  $L$  do not have subscripts. The second interval is defined relative to an unrestricted “best” firm, implying an efficient set with a 100% inclusion level. In many policy settings such a simplified joint confidence interval might be desired (e.g., how big a spread in efficiency rates is there between the best transplant program and the rest). At times, the total dispersion in efficiency is more relevant than each unit’s individual deviation from the uncertain best. The intervals displayed above were computed for the utility dataset and model with satisfactory results. For example, the interval shown in (5.2) had limits of  $L=0$  and  $U=0.843$ , implying a very large spread in the firms’ relative efficiencies.

Another type of multiple comparison that can be more easily made with the Bayesian approach than with the frequentist is cross-group comparisons. Questions such as “what is the probability that teaching hospitals are more efficient than non-teaching hospitals?” and “is there a statistically significant efficiency difference between private and government-owned utilities?” can be easily addressed. Estimates can also be computed for comparisons such as the probability that all (or some percentage) of the firms in one group are more efficient than all (or some percentage) of the firms in another group.

The biggest advantage to the Bayesian approach is in the results in this section and in further extensions along these lines. Given the numerical nature of the approach, the types of probability statements which are computable is limited only by research creativity and programming ability. In the frequentist approach, one must be able to derive the distribution needed to compute (simulate) the necessary critical values. This requirement to resolve for appropriate critical values adds considerable time and effort to the process and in some cases may defeat the less statistically oriented researcher.

## 6. 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 comparing the precision of the estimated efficiency scores and the resulting rankings of the firms studied. Horrace and Schmidt (2000) introduced to this literature two multiple comparison techniques (MCC and MCB) based on sampling theory statistics. In this paper, we add a Bayesian approach to the toolkit for measuring the precision of efficiency estimates and the ability of such estimates to accurately differentiate between the units being ranked.

After providing the details of how to implement our Bayesian 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 provides an empirical ranking of the 43 firms. Application of our Bayesian approach then allows us to analyze which firms can truly be differentiated from which others at any desired level of probability. The MCB and MCC approaches of Horrace and Schmidt and a parametric bootstrap approach were also applied to the same technical efficiency estimates. Differences among results were often substantial on average, due to the fundamental differences among the three methods. Typically, coverage was too broad with the frequentist method and too narrow with the parametric bootstrap method.

We believe that the Bayesian results provide more flexibility in terms of multiple comparisons that are possible. 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, generalizing this approach to compute such comparisons is considerably more complex. 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 typical (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 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 our Bayesian 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.

### **Acknowledgements**

The authors wish to thank Arnold Zellner, Gary Koop, John Geweke, Rolf Färe, Peter Schmidt, Robin Sickles, Bill Griffiths, Chris Cornwell, Robert Town, and

audience members at the 2002 American Agricultural Economics Association meetings in Long Beach, the 2002 European Productivity Workshop in Oviedo, Spain, 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.

## Notes

1. This is not an entirely new idea to Bayesians in this field. Fernandez et al. (2002) discuss the ability of Bayesian econometric approaches to easily allow inter-firm comparisons and plot several entire firm-specific efficiency measure posterior pdfs on a single set of axes to allow a visual comparison across firms. However, multiple comparisons are largely missing from the Bayesian efficiency estimation literature. For example, Koop et al. (1997), develop Bayesian tools for making inferences about firm-specific inefficiencies in panel data models, but concentrate on single-firm probability statements.
2. To make the following discussion simpler, particularly for non-Bayesians or new Bayesians, we will often use the term confidence interval in our discussion of Bayesian intervals that are more properly called posterior density regions.
3. A number of different specifications of input and output distance functions with bads could be estimated. For a discussion of the advantages of our adopted approach (see Atkinson and Dorfman, 2004).
4. Since the inclusion of  $v_{ft}$  makes the frontier distance function stochastic, it is possible for  $h(\epsilon_{ft})$  to be greater than 1.
5. Note that the time-varying nature of this specification is designed to improve the empirical relevance of the results. The basic nature of the problem is the same as if we had simply a firm-specific intercept model.
6. We are grateful to Professor Nelson for making his data available to us.
7. We employ three moment restrictions that involve both the structural parameters and the covariance matrix of the errors (see Atkinson and Dorfman 2004 Equation (4.7) for details).
8. On average 99% of our data points actually satisfied monotonicity for the inputs and outputs while 92% satisfied monotonicity for the bad.
9. Our Gibbs sampler relies on GMM estimates within every iteration to determine the conditional mean of  $\gamma$ .
10. We do not compute the bias-corrected and accelerated bootstrap confidence intervals as described in Efron and Tibshirani (1993), often termed  $BC_a$ . Because we are drawing sets of structural parameters from a multivariate normal distribution rather than running repeated regressions with bootstrapped data sets, the corrections for bias and skewness that are involved in computing those intervals seem unlikely to be needed. Further, leaving the intervals unadjusted seemed to us to create a fairer comparison to the Bayesian results.
11. Note that the bootstrap and Bayesian MCB intervals were explicitly chosen to minimize this criterion subject to maintaining the probability content. By contrast, the Horrace and Schmidt intervals are symmetrical and thus do not have minimal interval length.

## Appendix A

Putting the Gibbs sampler used in this application into a “recipe” form, the process could be described as follows.

0. Obtain initial value for the sum of squared errors matrix,  $\Xi^{(0)}$ , either through conventional GMM estimation as  $\hat{\mathbf{e}}'\hat{\mathbf{e}}$ , 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  $\Omega^{(i)}$  from  $\text{IW}(\Xi^{(i)}, \nu)$ , where  $\nu = (n - k + m + 1)/2$ . (Draw system covariance matrix conditional on covariance estimate in 1.) See Appendix A for further details on this step.
2. Compute  $\gamma_m^{(i)} = \text{GMM}(y, X, Z|\Omega^{(i)})$ . (Compute GMM estimate conditional on  $\Omega^{(i)}$ ). This requires iterating until convergence using GMM with the covariance of the errors held constant at  $\Omega^{(i)}$ .
3. Compute  $\Psi_m^{(i)} = \text{cov}(\gamma_m^{(i)})$ . (Calculate estimated covariance of GMM coefficients.)
4. Compute  $\Psi_p^{(i)} = [H_o^{-1} + (\Psi_m^{(i)})^{-1}]^{-1}$ . (Combining prior variance with data variance to get posterior variance.)
5. Compute  $\gamma_p^{(i)} = \Psi_p^{(i)}[H_o^{-1}g_o + (\Psi_m^{(i)})^{-1}\gamma_m^{(i)}]$ . (Combine prior mean of coefficients with moment conditions through the maxent principle to get posterior mean.)
6. Draw  $\gamma^{(i)}$  from  $\text{MVN}(\gamma_p^{(i)}, \Psi_p^{(i)})$ . (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.

## References

- Atkinson, S. and D. Primont. (2002). "Measuring Productivity Growth, Technical Efficiency, Allocative Efficiency, and Returns to Scale Using Distance Functions." *Journal of Econometrics* 108, 203–225.
- Atkinson, S. E. and J. H. Dorfman. (2004). "Crediting Electric Utilities for Reducing Air Pollution: Bayesian Measurement of Productivity and Efficiency.", Working Paper, University of Georgia.
- Atkinson, S. E., and P. Wilson. (1995). "Comparing Mean Efficiency and Productivity Scores from Small Samples: A Bootstrap Methodology." *Journal of Productivity Analysis* 6, 137–52.
- Baltagi, B. H. and J. M. Griffin. (1998). "A General Index of Technical Change." *Journal of Political Economy* 96, 20–41.
- 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.
- Chib, S. (1995). "Marginal Likelihood from the Gibbs Output." *Journal of the American Statistical Association* 90, 1313–1321.
- Cornwell, C., P. Schmidt and R.C. Sickles. (1990). "Production Frontiers with Time Series Variation in Efficiency Levels." *Journal of Econometrics* 46, 185–200.
- Efron, B and R. J. Tibshirani. (1993). *An Introduction to the Bootstrap* New York: Chapman and Hall.
- Färe, R. and D. Primont. (1995). *Multi-Output Production and Duality: Theory and Applications* Boston: Kluwer Academic Publishers.

- Fernández, C., G. Koop and M. Steel. (2002). "Multiple-Output Production with Undesirable Outputs." *Journal of the American Statistical Association* 97, 432–442.
- Hahn, G. J. and R. W. Hendrickson. (1971). "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.
- Horrace, W. C. and P. Schmidt. (2000). "Multiple Comparisons with the Best, with Economic Applications." *Journal of Applied Econometrics* 15, 1–26.
- Kim, J. Y. (2002). "Limited Information Likelihood and Bayesian Analysis." *Journal of Econometrics* 107, 175–193.
- Kim, Y. and P. Schmidt. (2000). "A Review and Empirical Comparison of Bayesian and Classical Approaches to Inference on Efficiency Levels in Stochastic Frontier Models with Panel Data." *Journal of Productivity Analysis* 14, 91–118.
- Koop, G., J. Osiewalski and M. F. J. Steel. (1997). "Bayesian Efficiency Analysis Through Individual Effects: Hospital Cost Frontiers." *Journal of Econometrics* 76, 77–106.
- Krinsky, I. and A. L. Robb. (1986). "On Approximating the Statistical Properties of Elasticities." *Review of Economics and Statistics* 68, 715–719.
- Nelson, R. A. (1984). "Regulation, Capital Vintage, and Technical Change in the Electric Utility Industry." *Review of Economics and Statistics* 66, 59–69.
- Tierney, L. (1994). "Markov Chains for Exploring Posterior Distributions (with discussion)." *Annals of Statistics* 22, 1701–1762.
- Zellner, A. and J. Tobias. (2001). "Further Results on Bayesian Method of Moments Analysis of the Multiple Regression Model." *International Economic Review* 42, 121–140.