Journal of Productivity Analysis, 14, 91–118 (2000)

A Review and Empirical Comparison of Bayesian and Classical Approaches to Inference on Efficiency Levels in Stochastic Frontier Models with Panel Data

YANGSEON KIM
University of Hawaii

PETER SCHMIDT
Michigan State University

Abstract

This paper applies a large number of models to three previously-analyzed data sets, and compares the point estimates and confidence intervals for technical efficiency levels. Classical procedures include multiple comparisons with the best, based on the fixed effects estimates; a univariate version, marginal comparisons with the best; bootstrapping of the fixed effects estimates; and maximum likelihood given a distributional assumption. Bayesian procedures include a Bayesian version of the fixed effects model, and various Bayesian models with informative priors for efficiencies. We find that fixed effects models generally perform poorly; there is a large payoff to distributional assumptions for efficiencies. We do not find much difference between Bayesian and classical procedures, in the sense that the classical MLE based on a distributional assumption for efficiencies gives results that are rather similar to a Bayesian analysis with the corresponding prior.

Keywords: Stochastic space frontier, Bayesian, bootstrap, MCB

1. Introduction

This paper considers the problem of interval estimation of technical efficiency levels in stochastic frontier models with panel data. The phrase “interval estimation” indicates that we are interested not only in point estimates of the efficiency levels of the individual firms, but also in confidence intervals for the efficiency levels. Such confidence intervals are of interest in the usual, obvious sense: when we report any estimate, we want to know how precise it is. Thus, if we report that a firm is 80% efficient, we would like to be able to distinguish between cases where our uncertainty is small (for example when we can be relatively sure that the true efficiency level is between 78% and 82%) and cases when our uncertainty is large (for example when we can be relatively sure only that it is between 65% and 95%). These issues are of academic interest and also of potential policy relevance. In the U.K, Australia and other countries, regulators in industries such as electricity, gas and water are using technical efficiency estimates to set their regulatory policies (such as price caps or allowable rates of return). See, for example, Førsund and Kittelsen (1998), Kittelsen

(1999) and Waddams Price (1999). Information on the accuracy of these estimates could have a major impact on the regulators' decisions, since one would presumably want to put less reliance on estimated efficiency levels that are known to be imprecise.

A number of different techniques have been proposed in the literature to address this problem. Given a distributional assumption for technical inefficiency, maximum likelihood estimation was proposed by Pitt and Lee (1981). Battese and Coelli (1988) showed how to construct point estimates of technical efficiency for each firm, and Horrace and Schmidt (1996) showed how to construct confidence intervals for these efficiency levels. Without a distributional assumption for technical inefficiency, Schmidt and Sickles (1984) proposed fixed effects estimation, and the point estimation problem for efficiency levels was discussed by Schmidt and Sickles (1984) and Park and Simar (1994). Simar (1992) and Hall, Härdle and Simar (1993) suggested using bootstrapping to conduct inference on the efficiency levels. Horrace and Schmidt (1996, 1999) constructed confidence intervals using the theory of multiple comparisons with the best, and Kim and Schmidt (1999) proposed a related technique which they called marginal comparisons with the best. Bayesian methods have been suggested by Koop, Osiewalski and Steel (1997) and Osiewalski and Steel (1998). They propose a model with an uninformative prior for firm-specific intercepts that is intended to be similar to the classical fixed effects model, and also models with informative priors, which are comparable to classical models that assume a distribution for inefficiency.

These models have been applied to various data sets, but there has been no systematic attempt to compare them all on a common data set. In this paper, we apply these models to three previously-analyzed data sets and compare the results. The major emphasis of the paper is to try to understand the relationship between the assumptions underlying the various models and the empirical results. More specifically, we are interested in two types of questions. First, we wish to see how much is gained, in terms of tightness of the confidence intervals, by being willing to make a distributional assumption for technical inefficiency. This is phrased as a classical question, but from a Bayesian perspective the question is simply rephrased as seeing how much is gained, in terms of tightness of the posterior distribution, by imposing an informative prior. Second, we wish to compare the results from Bayesian and classical analyses, where we match as far as possible the strength of the assumptions underlying the analyses. We find large gains from distributional assumptions, and we do not find much difference between classical and Bayesian analyses that rely on assumptions of comparable strength.

2. The Model

We begin with the basic panel data stochastic frontier model:

$$y_{it} = \alpha + x_{it}'\beta + v_{it} - u_i, \quad i = 1, \ldots, N, t = 1, \ldots, T. \tag{1}$$

Here i indexes firms or productive units and t indexes time periods. y_{it} is the scalar dependent variable representing the logarithm of output for the i^{th} firm in period t, α is a scalar intercept, x_{it} is a $K \times 1$ vector of functions of inputs (e.g., in logarithms for the Cobb-Douglas specification), β is a $K \times 1$ vector of coefficients and v_{it} is an i.i.d. error term with zero mean and finite variance. The u_i satisfy $u_i \geq 0$, and $u_i > 0$ is an indication of technical inefficiency. Note that u_i is time-invariant. For a logarithmic specification such as this the technical efficiency of the i^{th} firm is defined as $r_i = \exp(-u_i)$, so technical inefficiency is $1 - r_i$. For small values of u_i, u_i is approximately equal to $1 - \exp(-u_i) = 1 - r_i$, so that u_i itself is sometimes used as a measure of technical inefficiency.

Now define $\alpha_i = \alpha - u_i$. With this definition, (1) becomes the standard panel data model with time-invariant individual effects:

$$y_{it} = \alpha_i + x_{it}'\beta + v_{it} \tag{2}$$

Obviously we have $\alpha_i \leq \alpha$ and $u_i = \alpha - \alpha_i$. As before, technical efficiency is $r_i = \exp(-u_i)$.

The previous discussion regards zero as the minimal possible value of u_i, and α as the maximal possible value of α_i, over any possible sample; that is, essentially, as $N \to \infty$. For some purposes, and especially when N is not large, it is also useful to consider the following representation. We write the intercepts α_i in ranked order as:

$$\alpha_{(1)} \leq \alpha_{(2)} \cdots \leq \alpha_{(N)}, \tag{3}$$

so that (N) is the index of the firm with the largest value of α_i among firms $i = 1, \ldots, N$. It is convenient to rank the u_i in the *opposite* order, as $u_{(N)} \leq \cdots \leq u_{(2)} \leq u_{(1)}$, so that $\alpha_{(i)} = \alpha - u_{(i)}$ for all i. Then obviously $\alpha_{(N)} = \alpha - u_{(N)}$, and firm (N) has the largest value of α_i or equivalently the smallest value of u_i among firms $i = 1, \ldots, N$. We will call this firm the *best* firm in the sample. In some methods we measure inefficiency relative to the best firm in the sample, and this corresponds to considering the relative efficiency measures:

$$u_i^* = u_i - u_{(N)} = \alpha_{(N)} - \alpha_i, r_i^* = \exp(-u_i^*). \tag{4}$$

3. Classical Statistical Procedures

In this section we will discuss classical statistical procedures. Bayesian procedures will be discussed in the next section. We will distinguish procedures that make an assumption about the distribution of u_i from those that do not.

3.1. Efficiency Measurement with a Distributional Assumption for Inefficiency

In this subsection we consider estimation under strong assumptions that are similar to those made in the cross-sectional case (Aigner, Lovell and Schmidt (1977)). We assume independence across firms (values of i). We assume that the explanatory variables x_{it} are strictly exogenous: $(x_{i1}, \ldots, x_{iT})$ is independent of $(v_{i1}, \ldots, v_{iT}, u_i)$. We assume that the v_{it} are i.i.d. as $N(0, \sigma_v^2)$ and are independent of u_i. Finally, we assume a specific distribution for u_i. The distributions most commonly considered in the literature have been the half-normal and the exponential. Other suggestions include the truncated normal (Stevenson

(1980)) and gamma (Greene (1990)) distributions. In this paper we will use the exponential distribution, primarily because it is easiest to handle in the Bayesian framework, and we want to be able to make comparisons across Bayesian and classical approaches based on the same distribution for u_i. Thus we assume the following density for u_i: $f(u_i) = \lambda^{-1}\exp(-u_i/\lambda)$, with $E(u_i) = \lambda$.

Of course, simplicity is not the most noble of reasons for choosing a particular distribution. If our interest were primarily in a serious empirical analysis, as opposed to a comparison of empirical methods, we would certainly want to consider alternative distributional assumptions and ask which best fits the data.

The preferred method of estimation is maximum likelihood estimation (MLE). The likelihood function has been derived by Aigner, Lovell and Schmidt (1977) for the cross-sectional case ($T = 1$), and is easily extended to the case of panel data as follows:

$$\begin{aligned}\ln L \;=\; & -\ln(\lambda) - [N(T-1)/2]\ln(2\pi\sigma_v^2) - (N/2)\ln(T) \\ & - [T/(2\sigma_v^2)]\sum_{i=1}^{N}\left\{\left[\sum_{t=1}^{T}\varepsilon_{it}^2/T\right] - [\bar{\varepsilon}_i + \sigma_v^2/(T\lambda)]\right\} \\ & + \sum_{i=1}^{N}\ln[1 - \Phi(T^{1/2}\bar{\varepsilon}_i/\sigma_v + \sigma_v T^{-1/2}/\lambda)] \qquad (5)\end{aligned}$$

where Φ denotes the cdf of the standard normal distribution, $\varepsilon_{it} = v_{it} - u_i = y_{it} - \alpha - x_{it}{}'\beta$, and $\bar{\varepsilon}_i = T^{-1}\sum_{t=1}^{T}\varepsilon_{it}$. The likelihood function is maximized numerically to obtain the MLE of the parameters $(\alpha, \beta, \lambda, \sigma_v^2)$. These estimates are consistent as $N \to \infty$ for fixed T. The implication of this is that N needs to be large for MLE to be appropriate, whereas large T is not required.

The model can also be regarded as the random effects model from the panel data literature, so that estimation by generalized least squares (GLS) is possible. GLS is consistent as $N \to \infty$ without the assumption of normality of the v_{it} and without an assumption of a specific distribution for the u_i. However, the model above differs from the standard random effects model because u_i does not have a mean of zero. Thus the GLS estimated intercept will be a consistent estimate of $\alpha - E(u_i)$ rather than of α. It is possible to adjust the intercept upward, by adding a consistent estimate of $E(u_i)$, but this requires an assumption about the distribution of u_i. As a result we will consider only the MLE.

Estimation of the model yields residuals $\hat{\varepsilon}_{it} = y_{it} - \hat{a} - x_{it}{}'\hat{\beta}$, which are naturally regarded as estimates of $\varepsilon_{it} = v_{it} - u_i$, whereas we are interested in estimating u_i itself. The usual solution to this problem, following Jondrow et al. (1982) and Battese and Coelli (1988), is to consider the distribution of u_i conditional on $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})$, evaluated at $\hat{\varepsilon_i}$. Jondrow et al. give this distribution for the cross-sectional case ($T = 1$), for the cases that u_i is half-normal or exponential, and suggest the point estimate $E(u_i|\varepsilon_i)$. Battese and Coelli (1988) give both $E(u_i|\varepsilon_i)$ and $E(r_i|\varepsilon_i)$ for the half-normal case with panel data ($T \geq 1$). Actually, when the v_{it} are i.i.d. normal, the distribution of u_i conditional on $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})$ is the same as the distribution of u_i conditional on $\bar{\varepsilon}_i$, regardless of the distribution of u_i. Therefore the results of Jondrow et al. can be extended to the panel data case just by replacing ε_i by $\bar{\varepsilon}_i$ and σ_v^2 by σ_v^2/T. From their Theorem 2, we find that the distribution of

u_i conditional on ε_i is $N(-\mu_i, \sigma_v^2/T)$ truncated at zero, where $\mu_i = \overline{\varepsilon}_i + \sigma_v^2/(T\lambda)$. We can therefore calculate confidence intervals using this distribution. This idea was suggested by Horrace and Schmidt (1996), who implemented it assuming the half-normal distribution for u_i, and the results in this paper differ only because a different distribution (exponential) is assumed. See also Bera and Sharma (1999) for a similar analysis.

Conducting inference on u_i using the distribution of u_i conditional on ε_i is not suggested as an authentic Bayesian procedure, but it obviously has a Bayesian flavor. The main difference between this distribution and a Bayesian posterior distribution is that it relies on asymptotics to ignore the effects of parameter estimation, whereas the uncertainty due to parameter estimation will figure into the Bayesian posterior. We might expect this difference not to matter very much when N is large, however.

3.2. Estimation without a Distributional Assumption for Inefficiency: Fixed Effects

We now discuss estimation without a distributional assumption for the u_i. This subsection will give a brief review of the point estimation problem based on the fixed-effects estimates, and the next two subsections will consider different ways of constructing confidence intervals for inefficiency based on these estimates.

Fixed effects estimation refers to the estimation of the panel data regression model (2), treating the α_i as fixed parameters. Because the α_i are treated as parameters, we do not need to make any distributional assumption about the inefficiencies; nor do we need to assume that they are uncorrelated with the x_{it} or the v_{it}. We still assume strict exogeneity of the regressors x_{it} in the sense that $(x_{i1}, \ldots, x_{iT})$ is independent of $(v_{i1}, \ldots, v_{iT})$. We also assume that the v_{it} have zero mean and constant variance σ_v^2, and are not autocorrelated. We do not need to assume a distribution for the v_{it}.

The fixed effects estimate $\hat{\beta}$, also called the *within* estimate, may be calculated by regressing $(y_{it} - \overline{y}_i)$ on $(x_{it} - \overline{x}_i)$, or equivalently by regressing y_{it} on x_{it} and a set of N dummy variables for firms. We then obtain $\hat{\alpha}_i = \overline{y}_i - \overline{x}_i{\prime}\hat{\beta}$; or equivalently the $\hat{\alpha}_i$ are the estimated coefficients of the dummy variables. This leads to the following expression for $\hat{\alpha}_i$:

$$\hat{\alpha}_i = \alpha_i + \overline{v}_i - \overline{x}_i(\hat{\beta} - \beta), \tag{6}$$

to which we will make reference later. The fixed effects estimate $\hat{\beta}$ is consistent as $NT \to \infty$ (i.e., as either N or T approaches infinity), and its variance is of order $[N(T-1)]^{-1}$. For a given firm (i), the estimated intercept $\hat{\alpha}_i$ is a consistent estimate of α_i as $T \to \infty$. Large T is needed for the term $\overline{v_i}$ in (6) to become negligible.

Schmidt and Sickles (1984) suggested the following estimates of technical inefficiency, based on the fixed effects estimates:

$$\hat{\alpha} = \max_{i=1,\ldots,N} \hat{\alpha}_i;\ \hat{u}_i^* = \hat{\alpha} - \hat{\alpha}_i,\ i = 1, \ldots, N. \tag{7}$$

Since these estimates clearly measure inefficiency relative to the firm estimated to be the best in the sample, they are naturally viewed as estimates of $\alpha_{(N)}$ and u_i^*, that is, of relative rather than absolute inefficiency. For fixed N, $\hat{\alpha}$ is a consistent estimate of $\alpha_{(N)}$ and $\hat{u}_i^*$ is a consistent estimate of u_i^* as $T \to \infty$. However, it is important to note that in finite samples (for small T) $\hat{\alpha}$ is likely to be biased upward, since $\hat{\alpha} \geq \hat{\alpha}_{(N)}$ and $E(\hat{\alpha}_{(N)}) = \alpha_{(N)}$, where $\hat{\alpha}_{(N)}$ is the estimated intercept for the *unknown* best firm. That is, the "max" operator in (7) induces upward bias, since the largest $\hat{\alpha}_i$ is more likely to contain positive estimation error than negative error. This bias is larger when N is larger and when the $\hat{\alpha}_i$ are estimated less precisely. The upward bias in $\hat{\alpha}$ induces an upward bias in the $\hat{u}_i^*$ and a downward bias in $\hat{r}_i^* = \exp(-\hat{u}_i^*)$; we underestimate efficiency because we overestimate the level of the frontier.

Schmidt and Sickles argued that $\hat{\alpha}$ and the $\hat{u}_i^*$ are consistent estimates of α and the $\hat{u}_i$ if both N and T approach infinity; that is, if both N and T are large, we can regard the $\hat{u}_i^*$ as estimates of absolute and not just relative inefficiency. The argument is simple. As $T \to \infty$, $\hat{\alpha}$ and the $\hat{u}_i^*$ are consistent estimates of $\alpha_{(N)}$ and the u_i^*, as noted above. As $N \to \infty$, $u_{(N)}$ should converge to zero, so that $\alpha_{(N)}$ should converge to α and the u_i^* should converge to the corresponding u_i. A more rigorous treatment of the asymptotics for this model is given by Park and Simar (1994), who show that, in addition to $N \to \infty$ and $T \to \infty$, we need to require $T^{-1/2} \ln N \to 0$ in order to ensure the consistency of $\hat{\alpha}$ as an estimate of α. This latter requirement limits the rate at which N can grow relative to T, in order to ensure that the upward bias induced by the "max" operation disappears asymptotically.

3.3. Multiple and Marginal Comparisons with the Best

Multiple comparisons with the best (MCB) is a statistical technique that yields confidence intervals for differences in parameter values between all populations and the best population. Horrace and Schmidt (1996, 1999) have suggested its use to construct confidence intervals for the relative technical inefficiencies $u_i^* = \alpha_{(N)} - \alpha_i = u_i - u_{(N)}$ or $r_i^* = \exp(-u_i^*)$, which are indeed differences from the best.

Let $A = (\alpha_1, \ldots, \alpha_N)$ be the vector of intercepts for the N firms in the regression model (2). (It would be natural to refer to this vector as α, but that symbol has already been used for the intercept in model (1).) As before we denote $\max_{i=1,\ldots,N}\alpha_i$ by $\alpha_{(N)}$. Then MCB constructs a set S of possibly best populations, and a set of intervals (L_i, U_i), such that:

$$P[(N)\varepsilon S \text{ and } L_i \leq \alpha_{(N)} - \alpha_i \leq U_i \text{ for all } i] \geq 1 - c, \tag{8}$$

where $1 - c$ is a chosen confidence level (e.g. 0.95). (Again, it would be natural to use α in place of c for the tail probability, but the symbol α has already been used.) Thus with a given confidence level we have a set of populations that includes the best, and joint confidence intervals for all differences from the best. MCB was developed by Hsu (1981, 1984) and Edwards and Hsu (1983). A general exposition can be found in Hochberg and Tamhane (1987), Hsu (1996) and Horrace and Schmidt (1999).

To perform MCB, we need an estimate $\hat{A}$, distributed as $N(A, \sigma^2 C)$ with C known, and where either σ^2 is known, or we have an estimate σ^2, independent of $\hat{A}$, such that $\hat{\sigma}^2/\sigma^2$ is distributed as χ_v^2/v. In typical MCB applications to the efficiency measurement problem, $\hat{A}$ will come from the fixed-effects estimation of the panel data regression model (2), as discussed above, and there will be enough degrees of freedom that we can effectively take σ^2 as known. Normality of $\hat{A}$ requires either that the errors v_{it} are normal or that T is large enough for a central limit theorem to hold for the expressions in (6) above.

Standard MCB proceeds under the further assumption that $C = kI_N$ with k known. This assumption is usually motivated by discussion of the "balanced one way model" (e.g., Hsu (1996), p. 43) in which we have independent observations y_{it} $(i = 1, \ldots, N, t = 1, \ldots, T)$ distributed as $N(\alpha_i, \sigma^2)$. In this case $k = 1/T$. This is equivalent to the panel data regression model (2) if β were known, since then we have $(y_{it} - x_{it}\beta) = \alpha_i + v_{it}$, which is the balanced one-way model. Since standard MCB would be applicable if β were known, it is reasonable to presume that standard MCB is a good approximation if β is estimated sufficiently precisely. Recall from the previous subsection that the variance of $\hat{\alpha}_i$ is of order T^{-1}, while the variance of $\hat{\beta}$ is of order $[N(T-1)]^{-1}$. Thus it may generally be the case that standard MCB is approximately applicable when N is large. This point is discussed in more detail in Horrace and Schmidt (1999).

Now define the following notation. Let $E(1/2)$ be the $N - 1 \times N - 1$ correlation matrix with all correlations equal to $1/2$ (i.e., diagonal elements equal one, off-diagonal elements equal $1/2$). Let z be a multivariate random variable distributed as student-t with dimension $N - 1$, degrees of freedom v, and correlation matrix $E(1/2)$. Define $d^*(c)$ as the c-level critical value of $\max_{i=1,\ldots,N-1}|z_i|$; i.e., $P[\max_i|z_i| \le d^*(c)] = 1 - c$. Tabulations of $d^*(c)$ can be found in Hsu (1996) or Horrace (1998). Define $h(c) = d^*(c)(2k\hat{\sigma}^2)^{1/2}$, and define the set $S(c) = \{i \mid \hat{\alpha} \ge \max_{j=1,\ldots,N}\hat{\alpha}_j - h(c)\}$. Define L_i and U_i as follows:

$$L_i = \max[0, \min_{j\varepsilon S(c)}\hat{\alpha}_j - \hat{\alpha}_i - h(c)], U_i = \max[0, \max_{j\neq i}\hat{\alpha}_j - \hat{\alpha}_i + h(c)] \quad (9)$$

Then MCB provides the statement (8) above, with $S = S(c)$.

As noted above, standard MCB requires that the variance matrix of $\hat{A} = (\hat{\alpha}_1, \ldots, \hat{\alpha}_N)$ be proportional to an identity matrix; that is, the various $\hat{\alpha}_i$ are uncorrelated and have equal variance. *General* MCB allows this variance matrix to be of an arbitrary form. A discussion of general MCB can be found in Horrace and Schmidt (1999, section 3), and is too lengthy to include here. The empirical results in this paper are all obtained from general MCB, but standard MCB would have yielded similar results.

The MCB statement (8) is a multiple statement in the sense that the confidence intervals all hold jointly with (at least) the specified probability. We can also consider marginal (one at a time) confidence intervals, which are more directly comparable to the intervals provided by the other techniques we consider. Kim and Schmidt (1999) refer to this as marginal comparisons with the best, which we will abbreviate as MargCB. There are *standard* and *general* versions of MargCB, where the standard version makes the same assumption about the variance matrix of $\hat{A}$ as does standard MCB. We will discuss standard MargCB, but the empirical results of this paper use the general version. Let $t^*(c)$ be the two-sided c-level critical value of the (univariate) student-t distribution with v degrees of freedom; i.e., if z is distributed as student-t with v degrees of freedom, then $P[|z| \le t^*(c)] = 1 - c$. Define $g(c) = t^*(c)(2k\hat{\sigma}^2)^{1/2}$. Define the set $S(c)$ as above. Define L_i^m and U_i^m as follows:

$$L_i^m = \max[0, \min_{j\varepsilon S(c)}\hat{\alpha}_j - \hat{\alpha}_i - g(c/2)], U_i^m = \max[0, \max_{j\neq i}\hat{\alpha}_j - \hat{\alpha}_i + g(c/2)]. \quad (10)$$

Then Kim and Schmidt (1999) establish the following result:

$$P[(N)\varepsilon S(c) \text{ and } L_i^m \le \alpha_{(N)} - \alpha_i \le U_i^m] \ge 1 - c. \quad (11)$$

Both MCB and MargCB are conservative procedures. The events given in statements (8) and (11) hold with probability of at *least* $1 - c$, and the inequality occurs because of uncertainty about which firm is best. There will be considerable uncertainty about which firm is best when one or more of the α_i are nearly as large as $\alpha_{(N)}$ and when the $\hat{\alpha}_i$ have large sampling variance. In such cases the set $S(c)$ will be large and the confidence intervals will be wide. The other techniques discussed in this paper are not conservative and may be expected to yield narrower confidence intervals. However, those techniques rely on stronger assumptions and/or asymptotic theory and correspondingly may be more likely to yield inferences that are erroneous.

MCB is not designed as an asymptotic procedure. Indeed, the problem of comparing N populations is hard to conceptualize unless N is fixed. However, since econometricians often think in terms of asymptotics, the following comments may be helpful. First, as just noted, standard MCB may be approximately valid when N is large. Second, MCB assumes that the $\hat{\alpha}_i$ are normally distributed. This should be so if the errors v_{it} are normal *or* if T is large. Third, MCB establishes confidence intervals for the relative inefficiencies u_i^* or r_i^*, but if N is large these can also be regarded as confidence intervals for the absolute inefficiencies u_i or r_i.

3.4. Bootstrapping

We can use bootstrapping to construct confidence intervals for functions of the fixed effects estimates. The inefficiency measures $\hat{u}_i^*$ (as in equation (7)) and the efficiency measures $r_i^* = \exp(-\hat{u}_i^*)$ are functions of the fixed effects estimates and so bootstrapping can be used for inference on these measures.

Consider the general setting in which we have a parameter θ, and there is an estimate $\hat{\theta}$ based on a sample $z_1, \ldots, z_n$ of i.i.d. random variables. The estimator $\hat{\theta}$ is assumed to be regular enough so that $n^{1/2}(\hat{\theta} - \theta)$ is asymptotically normal. The following bootstrap procedure will be repeated many times, say for $b = 1, \ldots, B$ where B is large. For iteration b, construct pseudo data $z_1^{(b)}, \ldots, z_n^{(b)}$ by sampling randomly with replacement from the original data $z_1, \ldots, z_n$. From the pseudo data, construct the estimate $\hat{\theta}^{(b)}$. The basic result of the bootstrap is that, under fairly general circumstances, the asymptotic (large n) distribution of $n^{1/2}(\hat{\theta}^{(b)} - \hat{\theta})$ conditional on the sample is the same as the (unconditional) asymptotic distribution of $n^{1/2}(\hat{\theta} - \theta)$. Thus for large n the distribution of $\hat{\theta}$ around the unknown θ is the same as the bootstrap distribution of $\hat{\theta}^{(b)}$ around $\hat{\theta}$, which is revealed by a large number (B) of draws.

We now consider the application of the bootstrap to the specific case of the fixed effects estimates. Our discussion follows Simar (1992). Let the fixed effects estimates be $\hat{\beta}$ and $\hat{\alpha}_i$, from which we calculate $\hat{u}_i^*$ and $\hat{r}_i^* (i = 1, \ldots, N)$. Let the residuals be $\hat{v}_{it} = y_{it} - \hat{\alpha}_i - x_{it}\prime\hat{\beta} (i = 1, \ldots, N, t = 1, \ldots, T)$. The bootstrap samples will be drawn by resampling these residuals, because the v_{it} are the quantities analogous to the z's in the previous paragraph, in the sense that they are assumed to be i.i.d. and the $\hat{v}_{it}$ are the observable versions of the v_{it}. (The sample size n above corresponds to NT.) So, for bootstrap iteration $b(= 1, \ldots, B)$ we calculate the bootstrap sample $\hat{v}_{it}^{(b)}$ and the pseudo data $y_{it}^{(b)} = \hat{\alpha}_i + x_{it}\prime\hat{\beta} + \hat{v}_{it}^{(b)}$. From these data we get the bootstrap estimates $\hat{\beta}^{(b)}$, $\hat{\alpha}_i^{(b)}$, $\hat{u}_i^{*(b)}$ and $\hat{r}_i^{*(b)}$, and the bootstrap distribution of these estimates is used to make inferences about the parameters.

We note that the estimates $\hat{u}_i$ and $\hat{r}_i$ depend on the quantity $\max_{j=1,\ldots,N}\hat{\alpha}_j$. Since "max" is not a smooth function, it is not immediately apparent that this quantity is asymptotically normal, and if it were not the validity of the bootstrap would be in doubt. A rigorous proof of the validity of the bootstrap for this problem is given by Hall, Härdle and Simar (1995). They prove the equivalence of the following three statements: (i) $\max_{j=1,\ldots,N}\hat{\alpha}_j$ is asymptotically normal. (ii) The bootstrap is valid as $T \to \infty$ with N fixed. (iii) There are no ties for $\max_j\alpha_j$; that is, there is a unique index i such that $\alpha_i = \max_{j=1,\ldots,N}\alpha_j$. There are two important implications of this result. First, the bootstrap will not be reliable unless T is large. Second, this is especially true if there are near ties for $\max_j\alpha_j$, in other words, when there is substantial uncertainty about which firm is best.

We now turn to specific bootstrapping procedures, which differ in the way they draw inferences based on the bootstrap estimates. In each case, suppose that we are trying to construct a confidence interval for $u_i^* = \max_j\alpha_j - \alpha_i$. That is, for a given confidence level c, we seek lower and upper bounds L_i, U_i such that $P[L_i \le u_i^* \le U_i] = 1-c$. This statement should hold exactly for large T, and for small T it will be inaccurate to an unknown extent.

The simplest version of the bootstrap for the construction of confidence intervals is the *percentile bootstrap*. Here we simply take L_i and U_i to be the upper and lower $c/2$ fractiles of the bootstrap distribution of the $\hat{u}_i^{*(b)}$. More formally, let $\hat{F}$ be the bootstrap cumulative distribution function for $\hat{u}_i^*$, so that $\hat{F}(s) = P(\hat{u}_i^{*(b)} \le s) =$ the fraction of the B bootstrap replications in which $u_i^{*(b)} \le s$. Then we take $L_i = \hat{F}^{-1}(c/2)$ and $U_i = \hat{F}^{-1}(1-c/2)$.

The percentile bootstrap intervals are accurate for large T but may be inaccurate for small to moderate T. This is a general statement, but in the present context there is a more specific reason to be worried, which is the finite sample upward bias in $\max_j\hat{\alpha_j}$ as an estimate of $\max_j\alpha_j$. This will be reflected in improper centering of the intervals and therefore inaccurate coverage probabilities. Simulation evidence on the severity of this problem is given by Hall, Härdle and Simar (1993) and Kim (1999). Several more sophisticated (or at least more complicated) versions of the bootstrap have been suggested to construct more accurate confidence intervals. Hall, Härdle and Simar (1993, 1995) suggested the *iterated bootstrap*, also called the double bootstrap, which consists of two stages. The first stage is the usual percentile bootstrap, which constructs, for any given c, a confidence interval that is intended to hold with probability $1-c$. We will call these "nominal" $1-c$ confidence intervals. The second stage of the bootstrap is used to estimate the true coverage probability of the nominal $1-c$ confidence intervals, as a function of c. That is, if we define the function $\pi(c) =$ true coverage probability level of the nominal$1-c$ level confidence interval from the percentile bootstrap, then we attempt to evaluate the function $\pi(c)$. When we have done so, we find c^*, say, such that $\pi(c^*) = 1-c$, and then we use as our confidence interval the nominal $1-c^*$ level interval from the first stage percentile bootstrap, which we "expect" to have a true coverage probability of $1-c$.

The mechanics of the iterated bootstrap are uncomplicated but time-consuming. For each of the original (first stage) bootstrap iterations B, the second stage involves a set of B_2 draws from the bootstrap residuals, construction of pseudo data, and construction of percentile confidence intervals, which then either do or do not cover the estimate $\hat{\theta}$. The coverage probability function $\pi(c)$ is the fraction of times coverage occurs. Generally we take $B_2 = B$, so that the total number of draws has increased from B to B^2 by going to the iterated bootstrap. Theoretically, the error in the percentile bootstrap is of order $n^{-1/2}$ while the error in the iterated bootstrap is of order n^{-1}. There is no clear connection between this statement and the question of how well finite sample bias is handled.

An objection to the iterated bootstrap is that it does not explicitly handle bias. If the nominal 90% confidence intervals only cover 75% of the time, it simply insists on a higher nominal confidence level, like 98%, so as to get 90% coverage. That is, it just makes the intervals wider, when bias might more reasonably be handled by recentering the intervals. A technique that does recenter the intervals is the *bias-adjusted bootstrap* of Efron (1982, 1985). As above, let θ be the parameter of interest, $\hat{\theta}$ the sample estimate and $\hat{\theta}^{(b)}$ the bootstrap estimate (for $b = 1, \ldots, B$), and let $\hat{F}$ be the bootstrap cdf. For n large enough that the bootstrap is accurate, we should expect $\hat{F}(\hat{\theta}) = 0.5$, and failure of this to occur is a suggestion of bias. Now define $z_0 = \Phi^{-1}[\hat{F}(\hat{\theta})]$, where Φ is the standard normal cdf, and where $\hat{F}(\hat{\theta}) = 0.5$ would imply $z_0 = 0$. Let $z_{c/2}$ be the usual normal critical value; e.g. for $c = 0.05$, $z_{c/2} = z_{.025} = 1.96$. Then the bias-adjusted bootstrap confidence interval is $[L_i, U_i]$ with:

$$L_i = \hat{F}^{-1}[\Phi(2z_0 - z_{c/2})], U_i = \hat{F}^{-1}[\Phi(2z_0 + z_{c/2)}]. \tag{12}$$

For example, suppose that there is upward bias, reflected by the fact that 60% of the bootstrap draws are larger than $\hat{\theta}$, so that $\hat{F}(\hat{\theta}) = 0.4$. Then $z_0 = -0.253$, and for $c = 0.05$ we have $\Phi(2z_0 - z_{c/2}) = \Phi(-2.466) = 0.0068$ and $\Phi(2z_0 + z_{c/2}) = 0.937$. Thus our confidence interval comes from the lower tail 0.0068 fractile and the upper tail 0.063 fractile, and we have compensated for upward bias by moving the interval left. This seems intuitively reasonable.

The assumption that justifies the bias-adjusted bootstrap is that, for some monotone increasing function g, $g(\hat{\theta}) - g(\theta)$ is distributed as $N(-z_0\sigma, \sigma^2)$ and $g(\hat{\theta}^{(b)}) - g(\hat{\theta})$ is also distributed as $N(-z_0\sigma, \sigma^2)$, for some z_0, σ^2. (The first distribution is from the probability law of the sample and the second is the bootstrap probability distribution induced by resampling from the given sample.) Thus we have normality, and also equal biases and variances, for some transformation of θ. The transformation function g need not be known. This is an advantage in implementation, but a disadvantage in trying to decide whether the assumption holds. It is not known whether the bias-adjusted bootstrap is valid for our specific problem, but it performs relatively well in the simulations reported in Kim (1999).

The final version of the bootstrap that we will consider is the *bias-adjusted and accelerated bootstrap* (BCa) of Efron and Tibshirani (1993). This is intended to allow for the possibility that the variance of $\hat{\theta}$ depends on θ, so that a bias-adjustment also requires a change in variance. This correction depends on some quantities defined in terms of the so-called jackknife values of $\hat{\theta}$. For $i = 1, \ldots, n$, let $\hat{\theta}_{(i)}$ be the value of the estimate based on all observations other than observation i; and let $\hat{\theta}_{(\cdot)} = n^{-1}\sum_{i=1}^{n}\hat{\theta}_{(i)}$ be the average of these values. Then the "acceleration" factor a is defined by:

$$a = \sum_{i=1}^{n}(\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)})^3 / 6\left[\sum_{i=1}^{n}(\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)})^2\right]^{1.5} \tag{13}$$

With z_0 and $z_{c/2}$ defined as above, define

$$\begin{aligned} b_1 &= z_0 + (z_0 + z_{c/2})/[1 - a(z_0 + z_{c/2})], \\ b_2 &= z_0 + (z_0 + z_{1-c/2})/[1 - a(z_0 + z_{1-c/2})] \end{aligned} \tag{14}$$

Then the confidence interval is $[L_i, U_i]$, with $L_i = \hat{F}^{-1}[\Phi(b_1)]$ and $U_i = \hat{F}^{-1}[\Phi(b_2)]$. More discussion can be found in Efron and Tibshirani (1993, chapter 14).

4. Bayesian Procedures

In this section we discuss the Bayesian analysis of the stochastic frontier model. Bayesian analyses have been proposed and described in a series of papers by Koop, Osiewalski and Steel (hereafter KOS), especially KOS (1997) but also including van den Broeck, Koop, Osiewalski and Steel (1994), Koop, Steel and Osiewalski (1995), and Osiewalski and Steel (1998).

4.1. General Discussion

The basic Bayesian principles are straightforward. We have a set of observable data $Y \equiv (y_1, \ldots, y_n)$ and a vector θ (say of dimension k) of unobservable parameters. Let $p(\theta)$ be the prior density of θ and $p(Y|\theta)$ be the likelihood, where the prior is specified by the data analyst and the likelihood follows from the assumed model. Then Bayes Law says that:

$$p(\theta|Y) \propto p(\theta)p(Y|\theta) \tag{15}$$

where $p(\theta|Y)$ is the posterior density of θ and "$\propto$" indicates proportionality. The traditional interpretation is that both the prior and the posterior reflect subjective probability distributions of θ, one (the prior) prior to the observation of Y and the other (the posterior) after the observation of Y. Bayes Law shows how the subjective probability distribution of θ is modified by the observation of Y. The concept of subjective probability is controversial but Bayes Law itself is not, since it is just the usual rule for conditional probability.

Inference on the parameters is performed using the posterior distribution. Since θ is usually multidimensional, one must face the often considerable problem of obtaining the marginal posterior distribution for a single given parameter such as θ_i (for some specific value $1 \leq i \leq k$). The marginal posterior density of θ_i is in principle defined by integrating the joint posterior density of θ with respect to all elements of θ other than θ_i, but this integral may not be analytically tractable. An alternative is to make Monte Carlo draws from the posterior distribution $p(\theta|Y)$ and to use these to reveal whatever features of the distribution of θ_i are interesting. Some numerical problems related to such Monte Carlo procedures will be discussed below.

Bayesian methods treat the parameters as random and condition on the data, which is more or less exactly the opposite of what classical methods do. However, in the present context of the stochastic frontier model with panel data, these distinctions can become a bit blurred. The parameters will typically include α, β, σ_v^2, $u_1, \ldots, u_N$, and possibly some additional nuisance parameters. Existing classical treatments of this model have always treated α, β, and σ_v^2 as fixed, but the inefficiency terms $u_1, \ldots, u_N$ have often been treated as random and sometimes assigned a distribution. As discussed in section 3.1 above, inference on u_i is then performed using the distribution of u_i conditional on $(\varepsilon_{i1}, \ldots, \varepsilon_{iT})$, which is certainly similar to a posterior distribution. In fact this can be regarded as Bayesian inference in classical clothing. It differs from the Bayesian posterior in that it treats α, β and the nuisance parameters in the distributions of v_{it} and u_i as known, whereas the Bayesian posterior conditions only on the data. This difference is not likely to be substantial in practice because the parameters being taken as known are estimated based on NT observations (rather than just T observations for u_i) and should not contribute much variability to the Bayesian posterior.

4.2. The Bayesian Fixed Effects Model

In this section we discuss a model that KOS call the *standard individual effects* model (or SIE model). They regard it as one possible variant of the *Bayesian fixed effects* model, whereas we will just refer to it as *the* Bayesian fixed effects model, but this is only a semantic point.

This model postulates an "uninformative" prior for the basic parameters $\alpha_1, \ldots, \alpha_N$, β, σ_v^2: $p(\alpha_1, \ldots, \alpha_N, \beta, \sigma_v^2) \propto \sigma_v^{-2}$. (We do not regard this prior as uninformative, but again this is just a semantic point.) Note that, in contrast to random effects models to be discussed later, we do not attempt to identify α and $u_1, \ldots, u_N$ separately. Rather we simply measure relative inefficiency, by considering $u_i^* = \max_j \alpha_j - \alpha_i$ and $r_i^* = \exp(-u_i^*)$ as functions of the firm-specific intercepts. This is similar in spirit to the classical fixed-effects treatment.

The likelihood $p(Y \mid \alpha_1, \ldots, \alpha_N, \beta, \sigma_v^2)$ is the usual (classical) normal likelihood that would follow from treating the x_{it} as fixed and the v_{it} as i.i.d. normal. Specification of the prior and the likelihood defines the problem and implies the form of the posterior. The marginal posterior of $\alpha_1, \ldots, \alpha_N, \beta$ can be calculated analytically to be $(N+K)$-variate student t with $N(T-1) - K$ degrees of freedom. For any reasonable problem the number of degrees of freedom is large enough to treat the posterior as multivariate normal. The posterior mean of β is the classical fixed-effects estimate $\hat{\beta}$ as in section 3.2, and similarly the posterior mean of α_i is the fixed-effects estimate $\hat{\alpha}_i = \overline{y}_i - \overline{x}_i\hat{\beta}$. It is also the case that the posterior variance matrix for β and $\alpha_1, \ldots, \alpha_N$ is the same as the classical result for the variance matrix of $\hat{\beta}$ and $\hat{\alpha}_1, \ldots, \hat{\alpha}_N$. For all of these reasons the name Bayesian fixed effects model seems appropriate.

The posterior distribution of the inefficiency estimate u_i^* or r_i^* is potentially complicated, but is easily revealed by Monte Carlo draws from the multivariate normal posterior distribution of $\alpha_1, \ldots, \alpha_N$. For example, confidence intervals are easily constructed from the percentiles of these draws. These are the same confidence intervals that would be constructed by a classical econometrician via a simulation from the estimated distribution of $\hat{\alpha}_1, \ldots, \hat{\alpha}_N$. We suspect that they will also often be similar to the confidence intervals constructed by bootstrapping the fixed effects estimates. They will differ only to the extent that the empirical distribution of the residuals v_{it} is not similar to the distribution of i.i.d. normals (which would reflect a failure of the assumed model).

An important point, stressed by KOS (1997), is that the fixed effects model favors low efficiency. An uninformative (flat) prior for $\alpha_1, \ldots, \alpha_N$ implies an uninformative (flat) prior for u_i^*, but an informative prior for $r_i^* = \exp(-u_i^*)$. More specifically, if u_i^* has constant density on $[0, \infty)$, then $r_i^* = \exp(-u_i^*)$ has density proportional to r_i^{*-1} on (0, 1]. This is an improper (prior) density that loosely speaking puts infinitely more weight on low values of r_i^* than high ones; for any constant c in (0, 1), no matter how small, there is infinite weight on r_i^* in $(0, c)$ but finite weight on r_i^* in $(c, 1]$. One could argue about whether this reflects a problem with this specific prior, or with improper priors in general, but in any case it implies that we should expect the Bayesian fixed effects model to yield smaller posterior efficiencies than a model with more or less any (proper) informative prior. In a sense this fact is the Bayesian counterpart to the finite sample bias problem discussed in section 3.2. Whether treated in a classical or Bayesian way, the fixed effects model will tend to yield smaller efficiency values than models that assert a distribution for inefficiency.

4.3. Bayesian Random Effects Models

In this section we consider models that have an informative prior for u_i. This allows us to distinguish u_i from the overall intercept α, and so now we can estimate absolute inefficiency (u_i) instead of just relative inefficiency (u_i^*). Thus the parameters of the problem are $\alpha, \beta, \sigma_v^2, u_1, \ldots, u_N$ and ϕ, where ϕ (which is present only in some models) represents parameters in the distribution of u. In all cases we take the likelihood $p(Y|\theta)$ to be the same normal likelihood as in the fixed effects case. In all cases we use the uninformative prior for α, β and σ_v^2: $p(\alpha, \beta, \sigma_v^2) \propto \sigma_v^{-2}$. When ϕ does not exist, $u_1, \ldots, u_N$ is prior independent of $\alpha, \beta, \sigma_v^2$. When ϕ does exist, ϕ and $u_1, \ldots, u_N|\phi$ are prior independent of $\alpha, \beta, \sigma_v^2$.

The models we consider all assert in one way or another that u_i follows an exponential distribution with mean λ, so that $p(u_i|\lambda) = \lambda^{-1}\exp(-u_i/\lambda)$. They differ in how λ is treated.

The first model we consider postulates an uninformative prior for r_i. More precisely, the $r_i (i = 1, \ldots, N)$ are i.i.d. as uniform on (0, 1). This uninformative prior for r_i implies an informative prior for $u_i = -\ln\, r_i$; the u_i are i.i.d. with density proportional to $\exp(-u_i)$, so that u_i is exponential with $\lambda = 1$. For this model, because the value of λ is specified, there are no nuisance parameters (ϕ) in the distribution of u_i.

Our second model differs from the first only because it uses a different value of λ. The first model implied prior median efficiency of 0.5, which seems low for at least some applications. The second model chooses λ to imply prior median efficiency of 0.8. This is achieved by picking $\lambda = -\ln(0.8)/\ln(2) = 0.322$.

Our last two models differ from the first two in that λ is now treated as a parameter. That is, we specify a hierarchical prior in which conditional on λ the u_i are i.i.d. as exponential with mean parameter λ, and then we specify a prior for λ. Thus "ϕ" in the notation above now corresponds to λ, the nuisance parameter in the distribution of u_i. Whereas the u_i are mutually prior independent conditional on λ, their dependence on a common value of λ implies that unconditionally they are not prior independent. KOS (1997, p. 86) refer to this as the "common efficiency distribution" or CED model

For our third model, we specify an "uninformative" prior for λ: $p(\lambda) \propto \lambda^{-1}$. An interesting theoretical point is that this choice of prior does not lead to a proper posterior. This is shown by Fernandez, Osiewalski and Steel (1997, Proposition 1, p. 179).[1] Explicitly, in this case the integrating constant (or "predictive distribution") $p(Y) = \int p(Y|\theta)p(\theta)d\theta$ is not σ-finite; there exists a set of Y's with positive Lebesgue measure such that $p(Y)$ is infinite. In this case the posterior distribution arguably makes no sense, as a matter of principle. We will report empirical results for this choice of prior (and, interestingly, they seem very reasonable), but this is a "Bayesian" procedure that some or all Bayesians would regard as flawed.

For our fourth model, we follow KOS (1997) and assume that λ^{-1} is exponential with mean $-1/\ln(r_{\text{med}})$, where r_{med} is the specified prior median efficiency. We take $r_{\text{med}} = 0.8$ as above (whereas they used 0.85). As noted by KOS, this hierarchical prior implies that the prior distribution of u_i is three-parameter inverted beta, but this model differs from the model (which they call the MIED model) in which the u_i are i.i.d. as three-parameter inverted beta, because in the present model the u_i are dependent due to the common value of λ.

For each of the above models, the specification of the prior and of the form of the likelihood implies the form of the posterior, $p(\alpha, \beta, \sigma_v^2, u_1, \ldots, u_N, \phi|Y)$. In principle, inference on u_i would be conducted based on its marginal posterior, $p(u_i|Y)$, but the integrals needed to construct the marginal posterior analytically are intractable. Numerical integration techniques such as those used by van den Broeck et al. (1994) in the cross-sectional case are likely to be impractical in the present setting due to the dimensionality of the integral. We will follow Koop, Steel and Osiewalski (1995), KOS (1997) and Osiewalski and Steel (1998) in using Gibbs Sampling to make Monte Carlo draws from the joint posterior. These draws then reveal the posterior distribution of the parameters such as u_i; in particular, confidence intervals for u_i are easily constructed from the percentiles of the Monte Carlo draws. This is the same principle as was followed for the Bayesian fixed effects model, except that there Gibbs Sampling was unnecessary because of the simple form of the joint posterior.

Gibbs sampling is a general procedure that makes draws from a joint distribution by making iterated sequential draws from the conditional distributions. It is useful in cases like the present one in which the conditional distributions are much simpler than the joint distribution, so that we know how to make draws from them. We split the set of parameters into three subsets: $(\alpha, \beta, \sigma_v^2)$, $(u1, \ldots, u_N)$ and ϕ. Starting from some arbitrary starting values, say $(\alpha, \beta, \sigma_v^2)^{(0)}$, $(u_1, \ldots, u_N)^{(0)}$ and $\phi^{(0)}$, we generate random draws in sequence from the conditional distributions, and then we iterate. Thus, at step j, make the following draws:

$$(\alpha, \beta, \sigma_v^2)^{(j)} \text{ from } p(\alpha, \beta, \sigma_v^2|Y, (u_1, \ldots, u_N)^{(j-1)}, \phi^{(j-1)})$$

$$(u_1, \ldots, u_N)^{(j)} \text{ from } p(u_1, \ldots, u_N|Y, (\alpha, \beta, \sigma_v^2)^{(j-1)}, \phi^{(j-1)})$$

$$\phi^{(j)} \text{ from } p(\phi|Y, (\alpha, \beta, \sigma_v^2)^{(j-1)}, (u_1, \ldots, u_N)^{(j-1)}).$$

For large enough s that convergence has occurred, $(\alpha, \beta, \sigma_v^2)^{(s)}$, $(u_1, \ldots, u_N)^{(s)}$ and $\phi^{(s)}$ can be treated as draws from the joint posterior. The reader is referred to Dorfman (1997) for more discussion of Gibbs Sampling, and to KOS (1997, Appendix) for the forms of the conditional distributions needed in the present case.

5. Empirical Results

We now proceed to apply the classical and Bayesian procedures described above to three previously-analyzed data sets. These data sets were chosen to have rather different characteristics. The first data set consists of $N = 171$ Indonesian rice farms observed for $T = 6$ growing seasons. We have $\sigma_v^2 = 0.108$ and $\sigma_u^2 = 0.007$ (these values being the exponential MLE's). Inference on inefficiencies will be very imprecise because T is small and because σ_v^2 is large relative to σ_u^2. The second data set consists of $N = 10$ Texas utilities observed for $T = 18$ years, with $\sigma_v^2 = 0.003$ and $\sigma_u^2 = 0.020$. For this data set we can estimate inefficiencies much more precisely because T is larger and σ_v^2 is smaller relative to σ_u^2. The third data set consists of $N = 25$ Egyptian tileries observed for a maximum of $T = 22$ production periods, with $\sigma_v^2 = 0.113$ and $\sigma_u^2 = 0.057$. This is a case that is intermediate between the other two. We will see that the precision of estimation of the efficiency levels will indeed differ strikingly across these data sets, and that choice of technique will matter more where precision is low.

5.1. Indonesian Rice Farms

These data are due to Erwidodo (1990) and have been analyzed subsequently by Lee (1991), Lee and Schmidt (1993), Schmidt and Horrace (1996, 1999) and others. There are $N = 171$ rice farms and $T = 6$ six-month growing seasons. Output is rice in kilograms and inputs are land, labor, seed and two types of fertilizer. The functional form is Cobb-Douglas with some dummy variables added for region, seasonality, and some types of farming practices. For a complete discussion of the data see Erwidodo (1990).

Table 1 gives point estimates for the regression parameters for the fixed effects model (within estimates); for the classical MLE's based on the half-normal and exponential distributions for u; and for our four Bayesian models. For the Bayesian models the point estimates are the posterior means and the standard errors are the square roots of the posterior variances. Here and subsequently the heading "uninformative" refers to our third Bayesian model, with an uninformative prior for the exponential parameter; "hierarchical" refers to our fourth model, with an exponential prior for the inverse of the exponential parameter; and "$\lambda = .322$" and "$\lambda = 1$" refer to our second and first models, with exponential priors with specified values of λ, so as to imply median prior efficiencies of 0.8 and 0.5 respectively. The results in Table 1 are quite similar across techniques. If we were primarily interested in the regression parameters, as opposed to the firm-specific efficiency levels, it would really not make much difference which technique we picked.

Table 2 gives point estimates and 90% confidence intervals for the relative efficiency measures r_i^*, for the classical and Bayesian fixed effects models. For the classical fixed effects model, we give the usual point estimate based on the within estimates; the MCB and MargCB confidence intervals; and confidence intervals based on three versions of the bootstrap. We used the percentile method and the iterated (two stage) percentile method, and also the bias-adjusted and accelerated bootstrap (labelled BCa). The bias-adjusted bootstrap (without acceleration) gave very similar results to the bias-adjusted and accelerated bootstrap so we do not report them separately. We used $B = 1000$ bootstrap replications, except

Table 1. Estimates of parameters: Indonesian rice farm.

		RE Model					
		MLE		Bayesian			
Variables	FE Model	Half-normal	Exponential	Uninformative	Hierarchical	$\lambda = .322$	$\lambda = 1$
Constant		5.199	5.181	5.187	5.187	5.319	5.419
		(.194)	(.193)	(.193)	(.195)	(.206)	(.220)
Seed	.121	.134	.135	.135	.135	.129	.125
	(.030)	(.027)	(.027)	(.026)	(.027)	(.028)	(.029)
Urea	.092	.113	.113	.113	.113	.102	.096
	(.021)	(.018)	(.018)	(.018)	(.018)	(.019)	(.020)
TSP	.089	.076	.076	.076	.077	.084	.088
	(.013)	(.012)	(.011)	(.012)	(.011)	(.012)	(.012)
Labor	.243	.219	.217	.217	.217	.224	.234
	(.032)	(.029)	(.029)	(.029)	(.029)	(.031)	(.031)
Land	.452	.481	.483	.483	.483	.477	.466
	(.035)	(.031)	(.031)	(.030)	(.031)	(.033)	(.035)
DP	.034	.009	.008	.008	.008	.013	.020
	(.032)	(.029)	(.028)	(.029)	(.029)	(.029)	(.031)
DV1	.179	.176	.176	.176	.176	.177	.177
	(.041)	(.038)	(.038)	(.038)	(.039)	(.038)	(.040)
DV2	.175	.140	.136	.138	.138	.149	.161
	(.057)	(.052)	(.052)	(.052)	(.052)	(.054)	(.055)
DSS	.053	.049	.049	.050	.050	.050	.052
	(.022)	(.021)	(.021)	(.022)	(.022)	(.021)	(.021)
DR1		−.058	−.059	−.060	−.060	−.093	−.125
		(.049)	(.048)	(.049)	(.050)	(.072)	(.109)
DR2		−.047	−.045	−.046	−.046	−.075	−.112
		(.058)	(.057)	(.057)	(.059)	(.076)	(.105)
DR3		−.078	−.077	−.078	−.078	−.126	−.176
		(.062)	(.060)	(.062)	(.062)	(.080)	(.110)
DR4		.016	.021	.018	.020	−.004	−.025
		(.058)	(.056)	(.057)	(.059)	(.082)	(.117)
DR5		.082	.089	.088	.090	.092	.085
		(.060)	(.059)	(.059)	(.061)	(.079)	(.110)
σ_v^2	.108	.108	.108	.110	.110	.106	.106
		(.005)	(.005)	(.005)	(.005)	(.005)	(.005)
σ_u^2		.007	.007	.007	.006		
		(.003)					
λ			.081	.084	.089		
			(.019)	(.019)	(.018)		

that for the iterated bootstrap we used only $B = B2 = 200$ replications, to shorten the computational time. (For the other two data sets, which are smaller, we used $B = B2 = 1000$ replications for the iterated bootstrap.) For the Bayesian fixed-effects model, we report the posterior mean and the 90% confidence intervals based on the appropriate percentiles of the simulated posterior distribution.

There are 171 firms and so we report results only for a few of them. We report results for the three firms (164, 118 and 163) that are most efficient; for the firms at the 75th percentile (31), 50th percentile (15) and 25th percentile (16) of the efficiency distribution; and for the

Table 2. Estimates of efficiencies (fixed effect model): Indonesian rice farm.

		MCB				Bootstrap								
		Simultaneous		Marginal		Percentile method		BC_a		Iterated bootstrap		Bayesian SIE		
Firm No.	FE estimates	LB	UB	LB	UB	LB	UB	LB	UB	LB	UB	Mean	LB	UB
164	1	.583	1	.737	1	.748	1	.781	1	.712	1	.912	.717	1
118	.932	.508	1	.642	1	.681	1	.772	1	.643	1	.854	.657	1
163	.930	.509	1	.643	1	.674	1	.773	1	.642	1	.850	.653	1
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
31	.616	.328	1	.421	1	.441	.724	.529	.878	.420	.756	.565	.423	.731
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
15	.554	.300	1	.379	1	.398	.646	.471	.746	.379	.673	.509	.383	.656
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
16	.498	.266	1	.340	1	.358	.589	.423	.679	.342	.614	.458	.344	.590
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
117	.379	.203	1	.259	.974	.272	.445	.321	.517	.259	.465	.348	.258	.450
45	.365	.197	1	.250	.940	.262	.427	.310	.496	.249	.445	.336	.253	.432

two worst firms (117, 45). All of these rankings are according to the classical fixed effects estimates.

In terms of the point estimates of efficiency levels, the classical and Bayesian fixed effects estimates are relatively similar. The Bayesian estimates are a little lower, especially for the most efficient firms. The efficiency estimates overall are rather low, with median efficiency only a little over 0.5. This is as expected.

We next discuss the confidence intervals for efficiency levels. The MCB and MargCB intervals are quite wide, especially for the less efficient firms. In fact, they really are too wide to be of much use. The bootstrap and Bayesian intervals are narrower, but still disappointingly wide. The percentile method of the bootstrap and the Bayesian fixed effects model give intervals that are quite similar. As noted above, this is as expected. With only $T = 6$ observations per firm, the accuracy of the percentile bootstrap is suspect. No such statement applies to the Bayesian method, but with only $T = 6$ observations per firm, the prior is certainly not dominated by the data, and since the prior is arguably unreasonable, so are the posterior results. As expected, the iterated bootstrap intervals are wider than those from the percentile method, and the bias-adjusted and accelerated bootstrap confidence intervals are shifted to the right (in the direction of higher efficiency levels). Simulations reported in Kim (1999) suggest that the bias-adjustment is helpful, but with so few observations per firm all of the bootstrap methods are probably unreliable.

Table 3 gives the estimated efficiencies and the associated 90% confidence intervals for our random effects models, including the classical MLE's based on the half normal and exponential distributions and our four Bayesian models with informative priors. We will not discuss the half normal results other than to note that they are not too different from the exponential. All of the confidence intervals are disappointingly wide, as they were for the fixed effects models of Table 2. However, comparing Tables 2 and 3, it is apparent that the efficiency levels from the random effects models of Table 3 are considerably higher than those from the fixed effects models. As noted above, from the classical point of view this is a reflection of the bias in the fixed effects efficiency estimates, while from a Bayesian point of view it reflects the influence of the underlying prior which is very heavily weighted toward low efficiency levels.

The exponential MLE and the Bayesian model with an uninformative prior for the exponential parameter give strikingly similar results. It is not surprising that the results are similar but it is perhaps unexpected that they are as similar as they are. The Bayesian model with $\lambda = 1$ has a lower prior median efficiency (equal to 0.5) than the hierarchical model or the model with $\lambda = 0.322$ (both of which have prior median efficiency of 0.8) and correspondingly has lower posterior efficiencies. The choice of prior matters a fair amount, which is a reflection of the small amount of data ($T = 6$) per firm.

5.2. Texas Utilities

We next consider the Texas utility data of Kumbhakar (1996), which was also analyzed by Horrace and Schmidt (1996, 1999) and Kim and Schmidt (1999). There are $N = 10$ privately owned Texas electric utilities observed for $T = 18$ years. Kumbhakar estimated

Table 3. Estimates of efficiencies (random effect model): Indonesian rice farm.

	MLE						Bayesian											
	Half-normal			Exponential			Uninformative			Hierarchical			$\lambda = .322$			$\lambda = 1$		
Firm No.	Mean	LB	UB	Mean	LB	UB	Mean	LB	UB	Mean	LB	UB	Mean	LB	UB	Mean	LB	UB
164	.964	.903	.998	.973	.924	.999	.974	.923	.999	.973	.921	.999	.954	.871	.997	.931	.820	.995
118	.964	.902	.998	.973	.923	.999	.973	.921	.999	.972	.918	.998	.958	.881	.997	.936	.827	.996
163	.959	.889	.998	.970	.924	.999	.969	.912	.998	.968	.908	.998	.946	.851	.997	.912	.785	.993
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
31	.924	.823	.994	.949	.863	.997	.948	.854	.997	.946	.853	.997	.891	.750	.991	.828	.657	.975
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
15	.923	.792	.990	.935	.834	.996	.935	.823	.996	.931	.819	.995	.820	.655	.972	.696	.529	.884
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
16	.845	.725	.969	.886	.751	.990	.885	.729	.991	.879	.722	.988	.719	.560	.897	.611	.463	.780
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
117	.773	.658	.907	.792	.643	.956	.796	.623	.965	.785	.610	.958	.591	.463	.741	.496	.379	.633
45	.774	.659	.908	.795	.645	.957	.798	.621	.968	.788	.621	.958	.600	.473	.747	.517	.403	.652

Table 4. Estimates of parameters: Texas utilities.

		RE Model					
		MLE		Bayesian			
Variables	FE Model	Half-normal	Exponential	Uninformative	Hierarchical	$\lambda = .322$	$\lambda = 1$
Constant		−5.053	−5.064	−5.045	−5.048	−5.049	−5.060
		(.245)	(.245)	(.249)	(.254)	(.252)	(.254)
Labor	−.129	−.078	−.076	−.080	−.082	−.084	−.087
	(.046)	(.031)	(.029)	(.031)	(.032)	(.031)	(.032)
Capital	.628	.586	.585	.588	.590	.592	.600
	(.053)	(.049)	(.048)	(.049)	(.049)	(.049)	(.050)
Fuel	.565	.584	.584	.583	.582	.581	.579
	(.039)	(.038)	(.038)	(.038)	(.038)	(.039)	(.039)
σ_v^2	.003	.003	.003	.003	.003	.003	.003
		(.0003)	(.0003)	(.0003)	(.0003)	(.0003)	(.0003)
σ_u^2		.010	.020	.029	.030		
		(.005)					
λ			.143	.171	.173		
			(.048)	(.061)	(.061)		

a cost function, whereas we will estimate a Cobb-Douglas production function. Output is electric power generated and the inputs are measures of labor, capital and fuel. For more details on the data see Kumbhakar (1996).

Table 4 gives the regression parameter estimates. We will not comment on these except to note that the variance of the one-sided error is large relative to the variance of noise (e.g. 0.020 vs. 0.003 for the exponential MLE). This is the opposite of the case for the Indonesian rice farms. For this reason, and because T is larger here (18 vs. 6), we expect more precise estimates of efficiency levels and less sensitivity of the results to the choice of method for this data set than for the previous one.

Of course, it must be admitted that 18 years is a long time span for which to assume fixed efficiency levels. This is an inherent problem in models with fixed efficiency levels—if the time span is long enough to estimate them precisely, the assumption that they are fixed is questionable.

Table 5 gives the estimated efficiencies and 90% confidence intervals for the fixed effects models, while Table 6 gives the same results for the random effects models. The format is the same as for Tables 2 and 3, except that we can display results for all $N = 10$ firms. We can see the same patterns here as we did for the previous data set (though less distinctly since differences across techniques are smaller). The MCB and MargCB intervals are wider than the other intervals. The fixed effects models give lower estimates of efficiency levels than the random effects models. Comparable classical and Bayesian models give comparable results: the classical fixed effects results with the percentile bootstrap are quite similar to the Bayesian fixed effects results, and the results from the classical exponential MLE are quite similar to those from the Bayesian model with an uninformative prior for the exponential parameter. Bayesian models with higher prior efficiency levels yield higher posterior efficiency levels.

Table 5. Estimates of efficiencies (fixed effect model): Texas utilities.

		MCB				Bootstrap								
		Simultaneous		Marginal		Percentile method		BC_a		Iterated bootstrap		Bayesian SIE		
Firm No.	FE Estimates	LB	UB	LB	UB	LB	UB	LB	UB	LB	UB	Mean	LB	UB
5	1	.945	1	.960	1	.983	1	.983	1	.977	1	.997	.981	1
3	.916	.797	1	.806	1	.828	1	.832	1	.820	1	.914	.820	1
10	.861	.765	.984	.773	.979	.790	.926	.792	.927	.786	.928	.859	.784	.926
1	.835	.769	.984	.775	.976	.786	.876	.789	.878	.782	.878	.832	.783	.878
8	.820	.759	.973	.764	.964	.775	.858	.777	.860	.772	.860	.817	.771	.859
9	.806	.755	.972	.759	.961	.767	.840	.770	.843	.765	.842	.804	.766	.841
2	.801	.735	.942	.740	.934	.752	.843	.754	.845	.749	.845	.798	.748	844
7	.786	.716	.920	.722	.913	.735	.831	.736	.832	.730	.834	.784	.731	.832
6	.786	.715	.917	.720	.910	.732	.832	.734	.833	.730	.834	.784	.729	.833
4	.762	.707	.910	.711	.901	.721	.798	.725	.802	.719	.800	.760	.719	.799

Table 6. Estimates of efficiencies (random effect model): Texas utilities.

	MLE						Bayesian											
	Half-normal			Exponential			Uninformative			Hierarchical			$\lambda = .322$			$\lambda = 1$		
Firm No.	Mean	LB	UB	Mean	LB	UB	Mean	LB	UB	Mean	LB	UB	Mean	LB	UB	Mean	LB	UB
5	.987	.971	.999	.989	.974	.999	.982	.950	.999	.981	.947	.999	.979	.943	.999	.977	.940	.998
3	.978	.959	.996	.983	.964	.998	.969	.923	.998	.966	.921	.997	.964	.919	.996	.958	.913	.994
10	.908	.889	.927	.913	.894	.932	.902	.862	.937	.899	.859	.936	.896	.859	.931	.891	.854	.927
1	.864	.846	.882	.867	.849	.886	.858	.828	.887	.856	.825	.886	.853	.825	.880	.848	.820	.876
8	.846	.828	.864	.849	.832	.867	.841	.811	.870	.838	.808	.868	.835	.807	.862	.831	.803	.859
9	.826	.809	.843	.828	.811	.846	.821	.790	.851	.818	.788	.849	.815	.787	.844	.812	.783	.840
2	.831	.814	.848	.834	.817	.852	.826	.795	.857	.823	.792	.854	.820	.790	.848	.816	.787	.844
7	.817	.800	.834	.820	.803	.837	.811	.779	.842	.810	.776	.842	.806	.776	.836	.802	.772	.833
6	.820	.803	.837	.823	.806	.841	.814	.782	.845	.812	.781	.843	.809	.779	.838	.804	.776	.833
4	.786	.770	.801	.789	.772	.805	.781	.751	.810	.779	.749	.808	.776	.748	.803	.771	.744	.799

Table 7. Estimates of parameters: Egyptian tileries.

		RE Model					
		MLE		Bayesian			
Variables	FE Model	Half-normal	Exponential	Uninformative	Hierarchical	$\lambda = .322$	$\lambda = 1$
Constant		.856	.860	.871	.866	.916	1.031
		(.253)	(.246)	(.248)	(.250)	(.250)	(.276)
Labor	1.007	1.030	1.025	1.024	1.025	1.020	1.012
	(.040)	(.038)	(.038)	(.038)	(.038)	(.038)	(.039)
Machines	.042	.046	.046	.046	.046	.046	.045
	(.045)	(.033)	(.031)	(.031)	(.032)	(.033)	(.038)
σ_v^2	.115	.113	.113	.114	.114	.114	.115
		(.008)	(.007)	(.008)	(.008)	(.008)	(.008)
σ_u^2		.040	.057	.067	.065		
		(.014)					
λ			.239	.257	.254		
			(.056)	(.062)	(.060)		

However, we repeat that the main result of interest is the general comparison of the results for this data set with those from the previous data set. For this data set we can estimate efficiency levels precisely enough to make reasonable statements about them, and the choice of technique is not critically important. The most important aspect of the choice of technique is the choice of a fixed versus random effects model. The fixed effects models give lower efficiencies, and are suspect for the reasons discussed in sections 3.2 and 4.2 above. The gain from being willing to assert a distribution for inefficiency is large, and at least for this data set there is reasonable robustness to the choice of distribution.

An interesting question, but one on which we have no evidence, is whether we would still find robustness to the choice of distribution if we enlarged the set of distributions we considered. In our classical analyses, we considered only the exponential and half-normal distributions, both of which are one-parameter distributions with zero mode. In the Bayesian analyses, all of our priors were variants of the exponential distribution. Multiple-parameter distributions with non-zero mode, like Stevenson's (1980) truncated normal, might well make a larger difference. This would be worth trying in future research.

5.3. Egyptian Tileries

The last data set we consider is for Egyptian tileries. It was collected by Seale (1990) and has been analyzed by Horrace and Schmidt (1996, 1999) and Kim and Schmidt (1999). There are $N = 25$ small-scale Egyptian manufacturers of ceramic floor tiles, with observations for a maximum of $T = 22$ three-week production periods. There are some missing data points (production did not occur in some periods) and so this is an unbalanced panel. Output is square meters of tile, while inputs are labor and machine hours.

The results are given in Tables 7, 8 and 9. As with the Indonesian rice farms, we present results only for a subset of the firms, choosing firms at the same percentiles of the efficiency

Table 8. Estimates of efficiencies (fixed effect model): Egyptian tileries.

		MCB				Bootstrap								
		Simultaneous		Marginal		Percentile method		BC_a		Iterated bootstrap		Bayesian SIE		
Firm No.	FE Estimates	LB	UB	LB	UB	LB	UB	LB	UB	LB	UB	Mean	LB	UB
14	1	.739	1	.805	1	.782	1	.946	1	.754	1	.907	.783	1
24	.994	.724	1	.790	1	.775	1	.943	1	.752	1	.901	.775	1
25	.989	.724	1	.789	1	.771	1	.939	1	.745	1	.896	.769	1
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
18	.953	.670	1	.737	1	.734	1	.893	1	.708	1	.863	.732	1
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
4	.895	.648	1	.708	1	.692	.940	.850	1	.672	.967	.812	.688	.945
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
6	.645	.478	1	.518	.956	.505	.672	.619	.781	.491	.688	.585	.503	.671
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
7	.555	.416	.885	.451	.816	.436	.576	.536	.670	.423	.594	.502	.434	.577
8	.493	.358	.815	.392	.744	.380	.520	.469	.616	.368	.537	.447	.379	.521

Table 9. Estimates of efficiencies (random effect model): Egyptian tileries.

	MLE						Bayesian											
	Half-normal			Exponential			Uninformative			Hierarchical			$\lambda = .322$			$\lambda = 1$		
Firm No.	Mean	LB	UB	Mean	LB	UB	Mean	LB	UB	Mean	LB	UB	Mean	LB	UB	Mean	LB	UB
14	.943	.865	.995	.948	.874	.996	.944	.864	.996	.945	.862	.996	.936	.847	.994	.902	.787	.988
24	.950	.877	.996	.951	.879	.996	.947	.867	.996	.947	.869	.996	.939	.851	.995	.902	.789	.988
25	.942	.863	.995	.946	.870	.996	.944	.863	.996	.943	.859	.996	.935	.847	.995	.896	.782	.987
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
18	.929	.844	.993	.934	.852	.994	.928	.829	.994	.930	.834	.994	.918	.816	.991	.874	.752	.981
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
4	.885	.787	.978	.896	.799	.984	.891	.782	.986	.893	.782	.986	.876	.764	.977	.825	.704	.950
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
6	.675	.600	.755	.662	.588	.745	.660	.573	.752	.662	.577	.757	.644	.561	.735	.599	.515	.683
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
7	.582	.517	.651	.567	.504	.638	.565	.492	.644	.568	.495	.646	.552	.484	.625	.514	.443	.591
8	.540	.477	.607	.513	.454	.581	.511	.438	.592	.513	.439	.595	.497	.427	.574	.460	.388	.534

distribution as in section 5.1. We will not discuss these results in detail, but the same comparisons across techniques that held for the other two data sets hold here as well. Comparing results across data sets, the results here are more precise, and less dependent on choice of technique, than for the Indonesian rice farms; they are less precise, and more dependent on choice of technique, than for the Texas utilities. This is predictable because, both in terms of the sizes of N and T, and also in terms of the relative variances of noise and inefficiency, this data set has characteristics that are intermediate between those of the previous two data sets.

6. Concluding Remarks

In this paper we considered a large number of classical and Bayesian procedures to estimate technical efficiency levels of firms and to construct confidence intervals for these efficiency levels. We then applied these methods to three data sets with different characteristics that determine the difficulty of the estimation problem. Comparing results across data sets and across methods within a data set leads to some clear and important conclusions.

First, the estimation problem is easier when T is large and when the variance of noise is small relative to the variance of inefficiency, and harder when T is small and when the variance of noise is large relative to the variance of inefficiency. In easier problems we can estimate efficiency levels more precisely than in harder problems, and there is less sensitivity of the results to the choice of technique.

Second, we do not find much difference between classical and Bayesian methods if we match methods that depend on comparable assumptions. For example, the Bayesian fixed effects model gives similar results to those obtained by the percentile bootstrap applied to the fixed effects (within) estimates. As another example, the classical MLE based on the exponential distribution gives similar results to the Bayesian model in which the prior distribution for inefficiency is exponential, and there is an uninformative prior for the exponential parameter. Furthermore, the two approaches face a similar problem, in that the results may be “unreliable” when T is small. In the classical framework, “unreliable” means that asymptotically valid inference may not be valid in small samples, while in the Bayesian framework “unreliable” means that the prior will not be dominated by the data and so there is a lack of robustness to the choice of the prior. We do not mean to allege that these are the same problem, but simply that in either approach small T causes problems.

Third, the multiple comparisons with the best (MCB) and marginal comparisons with the best (MargCB) intervals are wider than any of the others we consider. This is not a desirable feature in a confidence interval, but on the other hand these intervals are valid for small T, and a conservative, valid interval at least can provide the correct message that in some cases we don’t know very much.

Finally, the main difference in results is between fixed effects and random effects models. Fixed effects models (either classical or Bayesian) yield much lower efficiency levels than random effects models, and there are good reasons to be skeptical of the fixed effects results. From a classical point of view, the fixed effects estimates of efficiency levels are biased downward; from the Bayesian point of view, the fixed effects model embodies a prior that is unreasonably heavily weighted toward low efficiency levels. Random effects models require a distributional assumption for inefficiency, which may be unattractive. However, making such an assumption yields large dividends in terms of precision of estimation *and* in terms of more reasonable (less downward biased) average levels of efficiency.

Acknowledgment

The second author gratefully acknowledges the financial suppoft of the National Science Foundation.

Notes

1. We were unaware of this theoretical issue when the first draft of the paper was written, and we thank Mark Steel for pointing it out to us.

References

Aigner, D. J., C. A. K. Lovell and P. Schmidt. (1977). "Formulation and Estimation of Stochastic Frontier Production Functions." *Journal of Econometrics* 6, 21–37.

Battese, G. E. and T. J. Coelli. (1988). "Prediction of Firm-Level Technical Efficiencies with a Generalized Frontier Production Function and Panel Data." *Journal of Econometrics* 38, 387–399.

Bera, A. and S. C. Sharma. (1999). "Estimating Production Uncertainty in Stochastic Frontier Production Function Models." *Journal of Productivity Analysis* 12, 187–210.

Broeck, J. van den, G. Koop, J. Osiewalski and M. Steel. (1994). "Stochastic Frontier Models: A Bayesian Perspective." *Journal of Econometrics* 61, 273–303.

Dorfman, J. H. (1997). *Bayesian Economics through Numerical Methods: A Guide to Econometrics and Decision-Making with Prior Information*. New York: Springer Verlag.

Edwards, D. G. and J. C. Hsu. (1983). "Multiple Comparisons with the Best Treatment." *Journal of the American Statistical Association* 78, 965–971. Corrigenda (1984), *Journal of the American Statistical Association* 79, 965.

Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans.* Philadelphia: Society for Industrial and Applied Mathematics.

Efron, B. (1985). "Bootstrap Confidence Intervals for a Class of Parametric Problems." *Biometrika* 72, 45–58.

Efron, B. and R. J. Tibshirani. (1993). *An Introduction to the Bootstrap*. New York: Chapman and Hall.

Erwidodo. (1990). "Panel Data Analysis on Farm-Level Efficiency, Input Demand and Output Supply of Rice Farming in West Java, Indonesia." unpublished dissertation, Department of Agricultural Economics, Michigan State University.

Fernandez, C., J. Osiewalski and M. F. J. Steel. (1997). "On the Use of Panel Data in Stochastic Frontier Models with Improper Priors." *Journal of Econometrics* 79, 169–193.

Førsund, F. R. and S. A. C. Kittelsen. (1998). "Productivity Development of Norwegian Electricity Distribution Utilities." *Resource and Energy Economics* 20, 207–224.

Greene, W. H. (1990). "A Gamma-Distributed Stochastic Frontier Model." *Journal of Econometrics* 46, 141–164.

Hall, P., W. Härdle and L. Simar. (1993). "On the Inconsistency of Bootstrap Distribution Estimators." *Computational Statistics and Data Analysis* 16, 11–18.

Hall, P., W. Härdle and L. Simar. (1995). "Iterated Bootstrap with Applications to Frontier Models." *Journal of Productivity Analysis* 6, 63–76.

Hochberg, Y. and A. C. Tamhane. (1987). *Multiple Comparison Procedures*. New York: John Wiley and Sons.

Horrace, W. C. (1998). "Tables of Percentages of the k-Variate Normal Distribution for Large Values of k." *Communications in Statistics: Simulation and Computation* 27, 823–831.

Horrace, W. C. and P. Schmidt. (1996). "Confidence Statements for Efficiency Estimates from Stochastic Frontier Models." *Journal of Productivity Analysis* 7, 257–282.

Horrace, W. C. and P. Schmidt. (1999). “Multiple Comparisons with the Best, with Economic Applications.” *Journal of Applied Econometrics*, forthcoming.

Hsu, J. C. (1981). “Simultaneous Confidence Intervals for All Distances from the Best.” *Annals of Statistics* 9, 1026–1034.

Hsu, J. C. (1984). “Constrained Simultaneous Confidence Intervals for Multiple Comparisons with the Best.” *Annals of Statistics* 12, 1145–1150.

Hsu, J. C. (1996). *Multiple Comparisons: Theory and Methods*. London: Chapman and Hall.

Jondrow, J., C. A. K. Lovell, I. Materov and P. Schmidt. (1982). “On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model.” *Journal of Econometrics* 19, 233–238.

Kim, Y. (1999). “A Study in Estimation and Inference on Firm Efficiency.” Unpublished dissertation, Department of Economics, Michigan State University.

Kim, Y. and P. Schmidt. (1999). “Marginal Comparisons with the Best and the Efficiency Measurement Problem.” Unpublished manuscript, Michigan State University.

Kittelsen, S. A. C. (1999). “Using DEA to Regulate Norwegian Electricity Distribution Utilities.” Presentation at the Sixth European Workshop on Efficiency and Producitivity Analysis, Copenhagen.

Koop, G., M. F. Steel and J. Osiewalski. (1995). “Posterior Analysis of Stochastic Frontier Models using Gibbs Sampling.” *Computational Statistics* 10, 353–373.

Koop, G., J. Osiewalski and M. F. Steel. (1997). “Bayesian Efficiency Analysis through Individual Effects: Hospital Cost Frontiers.” *Journal of Econometrics* 76, 77–106.

Kumbhakar, S. C. (1996). “Estimation of Cost Efficiency with Heteroscedasticity: An Application to Electric Utilities.” *Journal of the Royal Statistical Society, Series D* 45, 319–335.

Lee, Y. H. (1991). “Panel Data Models with Multiplicative Individual and Time Effects: Applications to Compensation and Frontier Production Functions.” Unpublished dissertation, Department of Economics, Michigan State University.

Lee, Y. H. and P. Schmidt. (1993). “A Production Frontier Model with Flexible Temporal Variation in Technical Efficiency.” *The Measurement of Productive Efficiency*, (Eds. H. O. Fried, C. A. K. Lovell and S.S. Schmidt). New York: Oxford University Press.

Osiewalski, J. and M. Steel. (1998). “Numerical Tools for the Bayesian Analysis of Stochastic Frontier Models.” *Journal of Productivity Analysis* 10, 103–117.

Park, B. U. and L. Simar. (1994). “Efficient Semiparametric Estimation in a Stochastic Frontier Model.” *Journal of the American Statistical Association* 89, 929–936.

Pitt, M. M. and L. F. Lee. (1981). “The Measurement and Sources of Technical Inefficiency in the Indonesian Weaving Industry.” *Journal of Development Economics* 9, 43–64.

Seale J. L. (1990). “Estimating Stochastic Frontier Systems with Unbalanced Panel Data: The Case of Floor Tile Manufactories in Egypt.” *Journal of Applied Econometrics* 5, 59–79.

Schmidt P. and R. C. Sickles. (1984). “Production Frontiers and Panel Data.” *Journal of Business and Economic Statistics* 2, 367–374.

Simar, L. (1992). “Estimating Efficiencies from Frontier Models with Panel Data: A Comparison of Parametric, Non-Parametric and Semi-Parametric Methods with Bootstrapping.” *Journal of Productivity Analysis* 3, 171–203.

Stevenson, R. E. (1980). “Likelihood Functions for Generalized Stochastic Frontier Estimation.” *Journal of Econometrics* 13, 57–66.

Waddams Price, C. (1999). “Efficiency and Productivity Studies in Incentive Regulation of UK Utilities.” Unpublished manuscript, Centre for Management under Regulation, University of Warwick.