On the Accuracy of Bootstrap Confidence Intervals for Efficiency Levels in Stochastic Frontier Models with Panel Data

Myungsup Kim
University of North Texas

Yangseon Kim
East-West Center

Peter Schmidt
Michigan State University

October 2006

Abstract

We study the construction of confidence intervals for efficiency levels of individual firms in stochastic frontier models with panel data. The focus is on bootstrapping and related methods. We start with a survey of various versions of the bootstrap. We also propose a simple parametric alternative in which one acts as if the identity of the best firm is known. Monte Carlo simulations indicate that the parametric method works better than the percentile bootstrap, but not as well as bootstrap methods that make bias corrections. All of these methods are valid only for large time-series sample size (T), and correspondingly none of the methods yields very accurate confidence intervals except when T is large enough that the identity of the best firm is clear. We also present empirical results for two well-known data sets.

JEL Classifications: C15, C23, D24

Keywords: Stochastic frontier, bootstrap, efficiency

The third author acknowledges support under a Marie Curie Transfer of Knowledge Fellowship of the European Community under contract MTKD-CT-014288.

1 Introduction

This paper is concerned with the construction of confidence intervals for efficiency levels of individual firms in stochastic frontier models with panel data. 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 efficiency, 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, 2000) constructed confidence intervals using the theory of multiple comparisons with the best, and Kim and Schmidt (1999) suggested a univariate version of comparisons with the best. Bayesian methods have been suggested by Koop, Osiewalski, and Steel (1997) and Osiewalski and Steel (1998).

In this paper we will focus on bootstrapping and some related procedures. We provide a survey of various versions of the bootstrap for construction of confidence intervals for efficiency levels. We also propose a simple alternative to the bootstrap that uses standard parametric methods, acting as if the identity of the best firm is known with certainty. We present Monte Carlo simulation evidence on the accuracy of the bootstrap and our simple alternative. Finally, we present some empirical results to indicate how these methods work in practice.

2 Fixed-Effects Estimation of the Model

Consider the basic panel data stochastic frontier model of Pitt and Lee (1981) and Schmidt and Sickles (1984),

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

where 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 $K \times 1$ column vector 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 time-invariant u_i satisfy $u_i \geq 0$, and $u_i > 0$ is an indication of technical inefficiency. The u_i are treated as "fixed," and no assumptions are made about them. For a logarithmic specification such as Cobb-Douglas, the technical efficiency of the i^{th} firm is defined as $r_i = \exp(-u_i)$, so technical inefficiency is $1 - r_i$.

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 $u_i = \alpha - \alpha_i$ and $\alpha_i \leq \alpha$ since $u_i \geq 0$. 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 $N \to \infty$. It is also useful to consider the following representation in a given sample size of N. We write the intercepts α_i in ranked order, as $\alpha_{(1)} \leq \alpha_{(2)} \leq \cdots \leq \alpha_{(N)}$ so that (N) is the index of the firm with largest value of α_i among N firms. It is convenient to write the values of u_i in the opposite ranked order, as $u_{(N)} \leq \cdots \leq u_{(2)} \leq u_{(1)}$, so that $\alpha_{(i)} = \alpha - u_{(i)}$. 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 N firms. 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:

$$u_i^* = u_i - u_{(N)} = \alpha_{(N)} - \alpha_i, \quad r_i^* = \exp(-u_i^*). \tag{3}$$

Fixed effects estimation refers to the estimation of the panel data regression model (2), treating the α_i as fixed parameters. We assume strict exogeneity of the regressors x_{it}, in the sense that $(x_{i1}, x_{i2}, \cdots, x_{iT})'$ are independent of $(v_{i1}, v_{i2}, \cdots, v_{iT})'$. We also assume that the v_{it} are i.i.d. with zero mean and constant variance σ_v^2. We do not need to assume a distribution for the v_{it}.

Some of these assumptions could be weakened. For example, we could allow for autocorrelation by assuming that the vectors $v_i = (v_{i1}, v_{i2}, \cdots, v_{iT})'$ are i.i.d. The implication would be that in the bootstrap we would resample blocks corresponding to v_i rather than individual observations corresponding to v_{it}.

The fixed effects estimates $\hat{\beta}$, also called the *within* estimates, may be calculated by regressing $(y_{it} - \bar{y}_i)$ on $(x_{it} - \bar{x}_i)$, or equivalently by regressing y_{it} on x_{it} and a set of N dummy variables for firms. We then obtain $\hat{\alpha}_i = \bar{y}_i - \bar{x}_i'\hat{\beta}$, or equivalently the $\hat{\alpha}_i$ are the estimated coefficients of the dummy variables. The fixed effects estimate $\hat{\beta}$ is consistent as $NT \to \infty$. For a given firm i, the estimated intercept $\hat{\alpha}_i$ is a consistent estimate of α_i as $T \to \infty$.

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

$$\hat{\alpha} = \max_j \hat{\alpha}_j, \quad \hat{u}_i^* = \hat{\alpha} - \hat{\alpha}_i. \tag{4}$$

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.[1]

We define some further notation. We write the estimates $\hat{\alpha}_i$ in ranked order, as follows:

$$\hat{\alpha}_1 \leq \hat{\alpha}_2 \leq \cdots \leq \hat{\alpha}_{[N]}. \tag{5}$$

So $[N]$ is the index of the firm with the largest $\hat{\alpha}_i$, whereas (N) was the index of the firm with the largest α_i. These may not be the same. Note that $\hat{\alpha}$ as defined in (4) above is the same as $\hat{\alpha}_{[N]}$, but it may not be the same as $\hat{\alpha}_{(N)}$, the estimated α for the *unknown* best firm. Note that $\hat{\alpha}_{(N)}$ is well-defined, but it is not a feasible estimate because (N) is unknown.

As $T \to \infty$ with N fixed, $\hat{\alpha}$ is a consistent estimate of $\alpha_{(N)}$ and $\hat{u}_i^*$ is a consistent estimate of u_i^*. 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 $\mathrm{E}(\hat{\alpha}_{(N)}) = \alpha_{(N)}$.[2] That is, the "max" operator in (4) 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[3] 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.

The methods of this paper may be used also on a number of extended versions of this model. For example, we can have an unbalanced panel, in which the value of T varies of i. Similarly, we

[1] Note that, following the notation used in the frontiers literature, the "*" refers to relative as opposed to absolute efficiency. The same "*" is often used in the bootstrap literature to represent bootstrap draws. We will use "(b)" to represent bootstrap draws.

[2] $\mathrm{E}(\hat{\alpha}_i) = \alpha_i$ for each individual (fixed) value of i, including $i = (N)$. The same statement is not true for $i = [N]$ because $[N]$ is not fixed; it is a random outcome.

[3] The bias is larger when N is larger because, so long as the u_i are i.i.d. draws from some (unknown) distribution, we will be closer to having a tie for $\max_j \alpha_j$ when N is bigger.

could consider the fixed effects version of the time varying efficiency model of Cornwell, Schmidt, and Sickles (1990). In both cases the fixed effects estimates are still well defined.

3 Construction of Confidence Intervals by Bootstrapping

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

We begin with a brief discussion of bootstrapping in the general setting in which we have a parameter θ, and there is an estimate $\hat{\theta}$ based on a sample $z_1, \cdots, z_n$ 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 procedure will be repeated many times, say for $b = 1, \cdots, B$ where B is large. For iteration b, construct pseudo data $z_1^{(b)}, \cdots, z_n^{(b)}$ by sampling randomly with replacement from the original data $z_1, \cdots, z_n$. From the pseudo data, construct the estimate $\hat{\theta}^{(b)}$. The basic result of the bootstrap is that, under 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 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 $\hat{r}_i^*$ $(i = 1, \cdots, N)$. Let the residuals be $\hat{v}_{it} = y_{it} - \hat{\alpha}_i - x_{it}'\hat{\beta}$ $(i = 1, \cdots, N, t = 1, \cdots, T)$. 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 $(= 1, \cdots, B)$ we calculate the bootstrap sample $\hat{v}_{it}^{(b)}$ and the pseudo data $y_{it}^{(b)} = \hat{\alpha}_i + x_{it}'\hat{\beta} + \hat{v}_{it}^{(b)}$. 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.

Hall, Härdle, and Simar (1995) prove that the bootstrap is valid for this problem as $T \to \infty$ with N fixed. More discussion of this point will be given in Section 5.

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 significance level c, we seek lower and upper bounds L_i, U_i such that $\mathrm{P}(L_i \leq u_i^* \leq U_i) = 1 - c$.

The simplest version of the bootstrap 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 cumulative distribution function (cdf) for $\hat{u}_i^{*(b)}$ so that $\hat{F}(s) = \mathrm{P}(\hat{u}_i^{*(b)} \leq s) =$ the fraction of the B bootstrap replications in which $\hat{u}_i^{*(b)} \leq s$. Then, we take $L_i = \hat{F}^{-1}(c/2)$ and $U_i = \hat{F}^{-1}(1 - c/2)$.

The percentile bootstrap intervals should be 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 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 in Section 6 of this paper.

Several more sophisticated versions of the bootstrap have been suggested to construct confidence intervals with higher coverage probabilities. 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 of $1 - c$. 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 $1 - c^*$ confidence 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 original estimate $\hat{\theta}$. The coverage probability function $\pi(c)$ is estimated by the rate at which a nominal c-level interval based on the iterated bootstrap estimates covers the original estimate $\hat{\theta}$. 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. For example, if the nominal 90% confidence intervals only cover 75% of the bootstrap estimates in the first stage, it insists on a higher nominal confidence level, like 98%, so as to get 90% coverage. That is, it 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 $b = 1, \cdots, B$), and $\hat{F}$ 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 a 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 $c = 0.1$, $z_{c/2} = z_{0.05} = 1.645$. Then, the bias-adjusted bootstrap confidence interval is $[L_i, U_i]$ with:

$$L_i = \hat{F}^{-1}(\Phi(2z_0 - z_{c/2})), \quad U_i = \hat{F}^{-1}(\Phi(2z_0 + z_{c/2})). \tag{6}$$

A related technique is the *bias-adjusted and accelerated bootstrap* (BC_a) 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 $i = 1, \cdots, n$, let $\hat{\theta}_{(i)}$ be the value of the estimate based on all observations other than observation i; and let $\hat{\theta}_{(\bullet)} = n^{-1}\sum_{i=1}^{n} \hat{\theta}_{(i)}$ be the average of these values. Then the "acceleration" factor a is defined by:

$$a = \frac{\sum_{i=1}^{n}\left(\hat{\theta}_{(\bullet)} - \hat{\theta}_{(i)}\right)^3}{6\left(\sum_{i=1}^{n}\left(\hat{\theta}_{(\bullet)} - \hat{\theta}_{(i)}\right)^2\right)^{1.5}}. \tag{7}$$

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

$$b_1 = z_0 + \frac{\left(z_0 + z_{c/2}\right)}{\left(1 - a\left(z_0 + z_{c/2}\right)\right)}, \quad b_2 = z_0 + \frac{\left(z_0 - z_{c/2}\right)}{\left(1 - a\left(z_0 - z_{c/2}\right)\right)}. \tag{8}$$

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

It is important to note that there are cases in which the acceleration factor fails to be defined. This happens when all the jackknifed estimates are the same, which yields zero both for the numerator and for the denominator of the acceleration factor. For example, one firm could be so dominantly efficient in the industry that jackknifing the best firm (in our case, dropping one time dimensional observation) would not change the efficiency rank for the best firm. When the acceleration factor is not defined, we set it equal to zero (that is, we just use the bias-adjusted bootstrap).

Simar and Wilson (1998) discuss a bootstrap method with a different kind of bias correction, which we will call the *bias corrected percentile* method. As above, let $\hat{\theta}$ be the estimate based on the data, and let $\hat{\theta}^{(b)}$ be a bootstrap estimate, $b = 1, \cdots, B$. Define $\widehat{bias} = \bar{\hat{\theta}}^{boot} - \hat{\theta}$ where $\bar{\hat{\theta}}^{boot}$ is the average of the B bootstrap estimates. Now define the bias-corrected bootstrap values: $\tilde{\theta}^{(b)} = \hat{\theta}^{(b)} - 2\widehat{bias}$. Then simply apply the percentile method using the bias-corrected values $\tilde{\theta}^{(b)}$.

We can similarly define the bias-corrected point estimate. Define $\tilde{\theta} = \hat{\theta} - \widehat{bias} = 2\hat{\theta} - \bar{\hat{\theta}}^{boot}$. Then the mean of the $\tilde{\theta}^{(b)}$ equals $\tilde{\theta}$. The motivation is that on average the difference between $\hat{\theta}$ and θ is approximately the same as the difference between $\bar{\hat{\theta}}^{boot}$ and $\hat{\theta}$. So removing the bias once would center the bootstrap values on $\hat{\theta}$, and we need to remove bias twice (i.e. subtract two times $\widehat{bias}$) to get them cover θ.

It is also possible to apply the BC_a method to the bias-corrected bootstrap values $\tilde{\theta}^{(b)}$. This is discussed by Simar and Wilson (1998, p.52) as an attempt to center the median of the $\tilde{\theta}^{(b)}$ on $\tilde{\theta}$. We will call this the *bias-corrected bootstrap with* BC_a.

Simar and Wilson (2000) discuss a slightly different type of bias-corrected bootstrap, which we will call *Hall's percentile method*. Let $-a$ be the $1 - c/2$ percentile of $(\hat{\theta}^{(b)} - \hat{\theta})$, and $-b$ be the $c/2$ percentile. Then the upper bound for the confidence interval for θ is $(\hat{\theta} + b)$ and the lower bound is $(\hat{\theta} + a)$. So the lower bound is: $2\hat{\theta} -$ (the $1 - c/2$ percentile of $\hat{\theta}^{(b)}$) and the upper bound is: $2\hat{\theta} -$ (the $c/2$ percentile of $\hat{\theta}^{(b)}$).

The relationship of Hall's percentile method to the bias-corrected percentile method is interesting. The lower and upper bounds for Hall's percentile method can be expressed in terms of the bias-corrected point estimate $\tilde{\theta}$, as follows:

$$\text{Lower bound } = \tilde{\theta} - [(\text{the } 1-c/2 \text{ percentile of } \hat{\theta}^{(b)}) - \bar{\hat{\theta}}^{(boot)}], \tag{9a}$$

$$\text{Upper bound } = \tilde{\theta} + [\bar{\hat{\theta}}^{(boot)} - (\text{the } c/2 \text{ percentile of } \hat{\theta}^{(b)})]. \tag{9b}$$

An alternative that should be very similar if the distributions are symmetric is:

$$\text{Lower bound } = \tilde{\theta} - [\bar{\hat{\theta}}^{(boot)} - (\text{the } c/2 \text{ percentile of } \hat{\theta}^{(b)})], \tag{10a}$$

$$\text{Upper bound } = \tilde{\theta} + [(\text{the } 1-c/2 \text{ percentile of } \hat{\theta}^{(b)}) - \bar{\hat{\theta}}^{(boot)}]. \tag{10b}$$

However, it is easy to show that these are exactly the same as the upper and lower bounds for the bias-corrected percentile method.

A final note is that the iterated bootstrap procedure can be applied to all of the methods discussed here, not just to the percentile bootstrap.

4 Direct Versus Indirect Intervals

The discussion of the previous section was presented in terms of confidence intervals for u_i^*. Now suppose instead that we are interested in confidence intervals for $r_i^* = \exp(-u_i^*)$. Here there are two possibilities. The first possibility is to use bootstrap methods where the parameter of interest ("θ" in the generic notation of the previous discussion) is r_i^*. We will call this the *direct method*. A second possibility is to construct a confidence interval for u_i^* and then translate it into a confidence interval for r_i^*. We will call this the *indirect method*. It is possible because r_i^* is a monotonic transformation of u_i^*. More specifically, if the confidence interval for u_i^* is $[L, U]$, the corresponding indirect method confidence interval for r_i^* is $[\exp(-U), \exp(-L)]$.

For some bootstrap methods, the direct and indirect methods yield the same result. This is true for the percentile bootstrap, and it is also true for the BC_a method so long as the acceleration factor is defined. However, for the bias-corrected percentile method (with or without BC_a) and for Hall's percentile method, the direct and indirect methods yield different results. This occurs because these methods use a bias correction that depends on averaging, and averaging is affected by the nonlinear transformation from u_i^* to r_i^*.

Our intuition suggests that the indirect method would work better in finite samples, because the estimate of u_i^* is more nearly a linear function of the data than the estimate of r_i^* is. We will present simulation evidence later that supports this intuition.

5 A Simple Alternative to the Bootstrap

In this section we propose a simple parametric alternative to the bootstrap. We begin with the following simple observation. We wish to construct a confidence interval for $u_i^* = \alpha_{(N)} - \alpha_i$, or $r_i^* = \exp(-u_i^*)$. If we knew which firm was best - that is, if we knew the index (N) - we could construct a parametric confidence interval for u_i^* of the form: $(\hat{\alpha}_{(N)} - \hat{\alpha}_i) \pm$ (critical value) $*$ (standard error), where "critical value" would be the appropriate $c/2$ level critical value of the standard normal distribution, and "standard error" would be the square root of the quantity: estimated variance of $\hat{\alpha}_{(N)}$ + estimated variance of $\hat{\alpha}_i$ − 2*estimated covariance of $(\hat{\alpha}_{(N)}, \hat{\alpha}_i)$. This interval would be valid asymptotically as $T \to \infty$ with N fixed. In fact, if the v_{it} are i.i.d. normal and we use the critical value from the student t distribution, this interval would be valid in finite samples as well.

Such a confidence interval is infeasible because the identity of the best firm is unknown. However, we can construct the confidence interval:

$$(\hat{\alpha}_{[N]} - \hat{\alpha}_i) \pm (\text{critical value}) * (\text{standard error}), \tag{11}$$

where as before $\max_j \hat{\alpha}_j = \hat{\alpha}_{[N]}$. That is, we use a confidence interval that would be appropriate if (N) were known, and we simply pretend that $[N] = (N)$. That is, we pretend that we do know the identity of the best firm. This is our "feasible parametric" confidence interval.

Two details should be noted. First, in calculating the standard error of $\hat{\alpha}_{[N]} - \hat{\alpha}_i$, we evaluate $\text{var}(\hat{\alpha}_{[N]})$ and $\text{cov}(\hat{\alpha}_{[N]}, \hat{\alpha}_i)$ using the standard formulas that ignore the fact that the index $[N]$ is data-determined. That is, again we pretend that $[N] = (N)$ is known. Second, although $\alpha_{(N)} - \alpha_i \geq 0$, the lower bound of the parametric confidence interval can be negative. If it is, we set it to zero. This corresponds to setting the upper bound of the relative efficiency measure $\hat{r}_i^*$ to one.

The feasible parametric confidence intervals are valid asymptotically, as $T \to \infty$ with N fixed. This is the same sense in which the bootstrap confidence intervals are valid asymptotically. To understand why this is true, we refer to the proof of the validity of the bootstrap for this problem given by Hall, Härdle, and Simar (1995). They prove the equivalence of the following three statements: (i) $\max_j \hat{\alpha}_j$ 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 (N) such that $\alpha_{(N)} = \max_j \alpha_j$. The point of the "no ties" requirement is to insure that $P([N] = (N)) \to 1$ as $T \to \infty$ (with N fixed). That is, as $T \to \infty$ we *do* know which is the best firm. For finite T, the intervals (11) are invalid because $[N]$ is not necessarily equal to (N), and therefore the intervals are improperly centered, and the standard errors they use are wrong. But both of these problems disappear in the limit (as $T \to \infty$ with N fixed).

There are two important implications of this discussion. First, neither the bootstrap nor the feasible parametric intervals will 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.

It may be useful to consider the fundamental sense in which the feasible parametric intervals differ from bootstrap intervals. As we have noted above, the "max" operator causes $\hat{\alpha}_{[N]}$ to be biased upward as an estimate of $\alpha_{(N)}$. This causes an upward bias in $\hat{u}_i^*$ and a downward bias in $\hat{r}_i^*$. We will call this "first-level bias." It disappears as $T \to \infty$. The percentile bootstrap also suffers from a "second-level bias", in the sense that $\mathrm{E}(\max_j \hat{\alpha}_j^{(b)}) > \hat{\alpha}_{[N]}$.[4] Obviously this occurs because the same "max" operator that is applied in the original sample is also applied in the bootstrap samples. As a result, we do not expect the percentile bootstrap to work well for this problem. The bias-corrected versions of the bootstrap use the size of the second-level bias (which is observable) to correct the first-level bias. The ability to do this is a significant advantage of these bootstrap methods.

6 Simulations

In this section we conduct Monte Carlo simulations to investigate the finite-sample reliability of confidence intervals based on bootstrapping and on the alternative procedure described in the last section. We are interested in the coverage rates of the confidence intervals and the way that they are related to bias in estimation of efficiency levels.

[4]This expectation is with respect to the distribution induced by the bootstrap resampling.

There are some other possible methods that could have been considered, but which we did not include in order to finish the simulations in finite time. These include multiple comparisons with the best (MCB) intervals as in Horrace and Schmidt (1996, 2000), marginal comparisons with the best intervals as in Kim and Schmidt (1999), and intervals based on the MLE as in Horrace and Schmidt (1996). Some results for these methods can be found in Kim (1999). Basically the intervals based on MLE are quite reliable if the distribution of inefficiency is correct, and the MCB intervals are exceedingly conservative.

The model is the basic panel data stochastic frontier model given in (1) above. However, we consider the model with no regressors so that we can concentrate our interest on the estimation of efficiencies without having to be concerned about the nature of the regressors. In an actual empirical setting, the regression parameters β are likely to be estimated so much more efficiently than the other parameters that treating them as known is not likely to make much difference.

Our data generating process is:

$$y_{it} = \alpha + v_{it} - u_i = \alpha_i + v_{it}, \quad i = 1, \cdots, N, \quad t = 1, \cdots, T, \tag{12}$$

in which the v_{it} are i.i.d. $N(0, \sigma_v^2)$ and the u_i are i.i.d. half-normal: that is, let $u_i = |\mathrm{u}_i|$ where $\mathrm{u}_i \sim N(0, \sigma_u^2)$. Since our point estimates and confidence intervals are based on the fixed effects estimates of $\alpha_1, \cdots, \alpha_N$, the distributional assumptions on v_{it} and u_i do not enter into the estimation procedure. They just define the data generation mechanism.

The parameter space is $(\alpha, \sigma_v^2, \sigma_u^2, N, T)$, but this can be reduced. Without loss of generality, we can fix α to any number, since a change in the constant term only shifts the estimated constant term by the same amount, without any effect on the bias and variance of any of the estimates. For simplicity, we fix the constant term equal to one.

We need two parameters to characterize the variance structure of model. It is natural to think in terms of σ_v^2 and σ_u^2. Alternatively, recognizing that σ_u^2 is the variance of the untruncated normal from which u is derived, not the variance of u, we can think instead in terms of σ_v^2 and $\mathrm{var}(u)$, where $\mathrm{var}(u) = \sigma_u^2(\pi - 2)/\pi$. However, we obtain more readily interpretable results if we think instead in terms of the size of total variance and the relative allocation of total variance between v and u. The total variance is defined as $\sigma_\epsilon^2 = \sigma_v^2 + \mathrm{var}(u)$. Olson, Schmidt, and Waldman (1980) used $\lambda = \sigma_u/\sigma_v$ to represent the relative variance structure, so that their parametrization was in terms of σ_ϵ^2 and λ. Coelli (1995) used σ_ϵ^2 and either $\gamma = \sigma_u^2/(\sigma_v^2 + \sigma_u^2)$ or

$\gamma^* = \text{var}(u)/(\sigma_v^2 + \text{var}(u))$. The choice between these two parameters is a matter of convenience. We decided to use γ^* due to its ease of interpretation, so that we use the parameters σ_ϵ^2 and γ^*. The reason this is a convenient parametrization (compared to the "obvious" choice of σ_v^2 and σ_u^2) is that, following Olson, Schmidt, and Waldman (1980), one can show that comparisons among the various estimators are not affected by σ_ϵ^2. The effect of multiplying σ_ϵ^2 by a factor of k holding γ^* constant, is as follows.

1. constant term: bias change by a factor of $\sqrt{k}$ and variance changes by a factor of k,
2. σ_v^2 and σ_u^2: bias changes by a factor of k and variance changes by a factor of k^2,
3. γ^* (or γ or λ): bias and variance are unaffected.

We set σ_ϵ^2 at 0.25 arbitrarily, so that the only parameters left to consider are (γ^*, N, T). We consider three values for γ^*, to include a case in which the variance of v dominates, a case in which the variance of u dominates, and an intermediate case. We take $\gamma^* = 0.1$, 0.5, and 0.9 to represent the above three cases. With $\sigma_\epsilon^2 = 0.25$, σ_v^2, var(u), and σ_u^2 are determined as follows for each value of γ^*.

1. $\gamma^* = 0.1$: $\sigma_v^2 = 0.225$, $\text{var}(u) = 0.025$, $\sigma_u^2 = 0.069$,
2. $\gamma^* = 0.5$: $\sigma_v^2 = 0.125$, $\text{var}(u) = 0.125$, $\sigma_u^2 = 0.344$,
3. $\gamma^* = 0.9$: $\sigma_v^2 = 0.025$, $\text{var}(u) = 0.225$, $\sigma_u^2 = 0.619$.

Four values of N and T are considered. In order to investigate the effect of changing N, we fix $T = 10$ and consider N =10, 20, 50, and 100. Similarly, T is assigned the values of 10, 20, 50, and 100 while fixing $N = 10$. This is done for each different value of γ^*.

For each parameter configuration (T, γ^*, N), we perform $R = 1000$ replications of the experiment. We will investigate the coverage rates of nominal 90% confidence intervals. With $R = 1000$, the standard error for the estimated coverage rate is $\sqrt{(0.9)(0.1)/1000} = 0.00945$, so two standard errors equals about 0.019. Many of the coverage rates in the experiments will fail to fall within two standard deviations of 0.10.

For each replication, we calculate the estimate of α, $\hat{\alpha} = \hat{\alpha}_{[N]}$; the infeasible estimate of α, $\hat{\alpha}_{(N)}$; the efficiency measures $\hat{u}_i^* = \hat{\alpha} - \hat{\alpha}_i$ and $\hat{r}_i^* = \exp(-\hat{u}_i^*)$, for each $i = 1, \cdots, N$. We then calculate the following confidence intervals, for both $\hat{u}_i^*$ and $\hat{r}_i^*$: the feasible and infeasible parametric intervals of Section 5; and the intervals corresponding to the percentile bootstrap, the BC_a bootstrap, the bias-corrected bootstrap, the bias-corrected bootstrap with BC_a, and Hall's percentile method bootstrap. For all bootstrap methods, we calculated confidence intervals for r_i^* using both the direct and the indirect methods, as discussed in Section 4.

The bootstrap results were based on $B = 1000$ bootstrap replications.

An important limitation of the study is that we did not consider the iterated bootstrap. All of the bootstrap confidence intervals could potentially have had their coverage rates improved by iteration. However, the computational burden of the iterated bootstrap would be very considerable in a Monte Carlo setting. If we used $B_2 = 1000$ iterations in the second bootstrap stage, the experiment would have taken approximately 1000 times as long, and the execution time would literally have been measured in years.

We are primarily interested in the biases of the point estimates and the coverage rates of the confidence intervals. These biases and coverage rates are reported as averages over both the N firms (where relevant) and the R replications.

We begin the discussion of our results with Table 1. Table 1 displays the bias of the fixed effects estimates. The entries are easily understood in terms of the identity:

$$\mathrm{E}(\hat{\alpha} - \alpha) = \mathrm{E}(\hat{\alpha} - \alpha_{(N)}) + \mathrm{E}(\alpha_{(N)} - \alpha) \tag{13a}$$

or, equivalently

$$\mathrm{E}(\hat{u}_i^* - u_i) = \mathrm{E}(\hat{u}_i^* - u_i^*) + \mathrm{E}(u_i^* - u_i). \tag{13b}$$

Column (1) gives the bias of $\hat{\alpha}(= \hat{\alpha}_{[N]})$ as an estimate of α, or equivalently the bias of $\hat{u}_i^*$ as an estimate of u_i. It can be positive or negative, because it is the sum of a positive term and a negative term. Column (2) gives the bias of $\hat{\alpha}$ as an estimate of $\alpha_{(N)}$, or equivalently the bias of $\hat{u}_i^*$ as an estimate of u_i^*. This is the "first-level bias" of Section 5. It is always positive. That is, $\hat{\alpha}$ is biased upward as an estimate of $\alpha_{(N)}$, because of the "max" operation that defines $\hat{\alpha} = \max_j \hat{\alpha}_j$. This bias increases with N, but decreases when T and/or γ^* increases. It disappears as $T \to \infty$ or $\gamma^* \to 1$. Column (3) gives the mean of $(\alpha_{(N)} - \alpha)$, or equivalently of $(u_i^* - u_i)$. Its value is the negative of $u_{(N)} = \min_j u_j$, and correspondingly it is always negative. Its value does not depend on T, and it decreases (in absolute value) when N increases or γ^*

decreases.

Intuitively, column (2) shows that $\hat{u}_i^*$ overestimates relative inefficiency (u_i^*). Column (3) says that relative inefficiency is smaller than absolute inefficiency. Therefore, column (1) shows that $\hat{u}_i^*$ may overestimate or underestimate absolute inefficiency (u_i).

We now turn our attention to question of the accuracy of the various types of confidence intervals we have discussed. We present results for 90% confidence intervals for $r_i^* = \exp(-u_i^*)$. We are primarily interested in the coverage rates of the intervals, and the proportions of observations that fall below the lower bound and above the upper bound, but we also show the average width of the confidence intervals.

We begin with confidence intervals for r_i^* constructed using the indirect method. The coverage rates for these are exactly the same as the coverage rate for the confidence intervals for u_i^* from which they are derived. The reason we present intervals for r_i^* (rather than u_i^*) is that it is bounded between zero and one, and so the average width of the intervals is easier to interpret.

Table 2 gives the results for the *infeasible* parametric intervals of Section 5. The coverage rates of these intervals are very close to 0.90, as they should be. These intervals are infeasible in practice, since they depend on knowledge of the identity of the best firm, but they illustrate two points. First, for obvious reasons, the intervals are narrower when T is large and when γ^* is large (that is, when the variance of inefficiency is large relative to the variance of noise). The number of firms, N is not really relevant if we know which one is best. Second, and more importantly, there is no difficulty in constructing accurate confidence intervals for technical efficiency if we know which firm is best. All of the problems that we will see with the accuracy of feasible intervals are due to not knowing with certainty which firm is best.

Table 2 also gives the results for the feasible parametric method of Section 5 and for the percentile bootstrap. Consider first the percentile bootstrap. Its coverage rate is virtually always less than the nominal level of 90%, and sometimes it is very much less. The fundamental problem is that the intervals are not centered on the true values, due to the bias problem discussed above. (The upward bias of $\hat{\alpha}$ as an estiamte of $\alpha_{(N)}$ corresponds to an upward bias in $\hat{u}_i^*$ and a downward bias in $\hat{r}_i^*$. Thus too many r_i^* lie above the upper bound of the confidence intervals.) However, comparing the percentile bootstrap intervals to the infeasible parametric intervals, it is also the case that the percentile bootstrap intervals are too narrow.

Theoretically, the intervals should be accurate in the limit (as $T \to \infty$ with N fixed), and so the validity of the percentile bootstrap depends on large T. The bias problem is small when we have large T and γ^* and small N, and the coverage probability reaches almost 0.9 for these cases, but it falls in the opposite cases where the bias is big. The width of the intervals decreases as T or γ^* increases. However, the intervals get narrower with larger N, while the bias increases as N increases. This explains why the coverage probabilities of the percentile intervals fall very rapidly as N increases.

Now consider the feasible parametric intervals. These are clearly more accurate than the percentile bootstrap intervals. This is especially true in the worst cases. For example, for $T = 10$, $\gamma^* = 0.1$ and $N = 100$, compare coverage rates of 0.199 for the percentile bootstrap and 0.664 for the parametric intervals. The parametric intervals are wider and they are better centered, both of which imply higher coverage rates. To understand the point about better centering, recall the discussion of bias in Section 5. The parametric intervals have one level of bias ($\hat{\alpha}$ is a biased estimate of $\alpha_{(N)}$) whereas the percentile bootstrap has two ($\hat{\alpha}$ is a biased estimate of $\alpha_{(N)}$, and $\max_j \hat{\alpha}_j^{(b)}$ is a biased "estimator" of $\hat{\alpha}$).

Table 3 gives our results for the varieties of the bootstrap that make some sort of explicit bias adjustment for bias. Specifically, this includes the BC_a bootstrap, the bias-corrected bootstrap, the bias-corrected bootstrap with BC_a, and Hall's percentile method. As a general statement, these methods are fairly similar. None of them is clearly superior to the others.

A more interesting comparison is with the feasible parametric intervals. For definiteness we will compare the bias-corrected bootstrap and the feasible parametric intervals. The bias-corrected bootstrap is considerably better in the most difficult cases (where bias is largest), that is, when T is small, γ^* is small and N is large. The feasible parametric intervals have slightly better coverage rates in the easiest cases. This is true primarily because they are wider. Overall, the bias-corrected bootstrap seems a better bet than the feasible parametric method because it is so much better in the difficult cases and only slightly worse in the easy ones.

Table 4 gives some calculations of the various biases involved in this problem, which show why bias correction is useful. The entries are mean values of various quantities, with the average taken over replications and observations. Column (1) gives the mean value of $\alpha_{(N)}$. Column (2) gives the mean value of $\hat{\alpha}_{[N]}$, and column (3) gives the mean value of $\max_j \hat{\alpha}_j^{(b)}$. Thus, column

(4), which equals (2) − (1), gives "first-level bias" while column (5), which equals (3) − (2), gives "second-level bias." Column (6), which equals (4) − (5), we will call "remaining bias." It is the bias of the bias-corrected point estimate. The bias correction assumes that first-level bias and second-level bias are (approximately) the same, in which case remaining bias would be approximately zero.

Consider, for example, the first row of Table 4, corresponding to $T = 10$, $\gamma^* = 0.1$, $N = 10$. First-level bias equals 0.129, which means that the feasible parametric intervals (for u_i^*) would be mis-centered by 0.129. Second-level bias equals 0.074, so the percentile bootstrap intervals would be mis-centered by $0.129 + 0.074 = 0.203$. Remaining bias equals $0.129 - 0.074 = 0.055$, so that the bias-corrected bootstrap intervals would be mis-centered by 0.055.

For most other parameter values all of the biases are smaller than those just cited. However, the general conclusion remains that bias-correction removes some, but not all, of the bias.

We now turn to a comparison of direct versus indirect intervals for r_i^*, as discussed in Section 4. Table 5 gives the results for the direct intervals for the methods that make bias adjustments, and can be compared to Table 3, which gives the same results for the indirect intervals. We do not consider direct intervals based on our parametric methods. We also do not display the results for the percentile bootstrap, since the coverage rates of the direct and indirect intervals would be exactly the same. Similarly, for the BC_a intervals, the coverage rates in Tables 3 and 5 are only very slightly different; these differences arise from the few cases in which the acceleration factor for BC_a is not defined.

For the other three methods (bias-corrected bootstrap, bias-corrected bootstrap with BC_a, and Hall's percentile method), the coverage rates of the indirect intervals are generally higher. This is especially true in the most difficult cases. The basic reason why the indirect intervals cover better is that they are wider. As noted above, the superiority of the indirect intervals is intuitively reasonable, since we are basing the intervals on a more nearly linear function of the data.

Our last set of simulations is designed to consider cases in which the identity of the best firm is clear. Here we set one u_i at the 0.05 quantile of the half normal distribution, while the other $(N - 1)$ are set at equally spaced points between the 0.75 and 0.95 quantiles, inclusive. These u_i are then held fixed across replications of the experiment. The only randomness therefore comes from the stochastic error v. Since the identity of the best firm should be clear, the bias caused by the max operator should be minimal.

Table 6 gives the bias of the fixed effects estimates, and is of the same format as Table 1. Column (2) gives $\mathrm{E}(\hat{\alpha} - \alpha_{(N)})$, which is the bias component caused by the max operator. It is indeed much smaller than in Table 1.[5]

Correspondingly, we expect the various bootstrap and parametric intervals to be more accurate in the current cases than in the previous ones. Comparing Tables 2 and 7, this is certainly the case for the feasible parametric method and the percentile bootstrap. These methods now work quite well in almost all cases. However, the feasible parametric intervals now sometimes have coverage rates that are too high, presumably because the intervals are too wide. Comparing Tables 3 and 8, for the bootstrap methods that make explicit bias adjustments, it is not clearly the case that they do better now that the identity of the best firm is clear. In fact, comparing Tables 7 and 8, the feasible parametric method and the percentile bootstrap now work as well or better than the BC_a bootstrap, the bias-corrected bootstrap, the bias-corrected bootstrap with BC_a, or Hall's percentile method. It is counterproductive to try to control for bias when there is little or no bias.

The overall conclusion we draw from our simulations are straightforward. If it is clear from the data which firm is best, all of the methods of constructing confidence intervals work fairly well. There is no need to consider more complicated procedures than the percentile bootstrap. The parametric intervals are also reliable, but they may be wider than necessary. Conversely, if the time series sample size T is not large enough for the identity of the best firm to be clear, none of the methods of constructing confidence intervals are very reliable. The percentile bootstrap is particularly bad. A bias-corrected method should be used in such cases.

7 Empirical Results

We now apply the procedures described above to two well-known 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. For this data set, the variance of stochastic

[5]There are a few small negative entries. This cannot be so, in principle, and these negative entries are due to the randomness of the experiment.

noise (v) is large relative to the variability in u (var(u)): that is, $\hat{\gamma}^* = 0.222$ with $\hat{\sigma}_\epsilon^2 = 0.138$. Inference on inefficiencies will be very imprecise because T is small, $\hat{\gamma}^*$ is small and N is large. The second data set consists of $N = 10$ Texas utilities observed for $T = 18$ years. For this data set, σ_v^2 is small relative to var(u): $\hat{\gamma}^* = 0.700$ with $\hat{\sigma}_\epsilon^2 = 0.010$. In this case we can estimate inefficiencies much more precisely because T and γ^* are larger, and N is smaller. We will see that the precision of the estimates will differ across these data sets, and that choice of technique matters more where precision is low. A more detailed analysis of these data, including Bayesian results and results for multiple and marginal comparisons with the best, can be found in Kim and Schmidt (1999).

7.1 Indonesian Rice Farms

These data are due to Erwidodo (1990) and have been analyzed subsequently by Lee (1991), Lee and Schmidt (1993), Horrace and Schmidt (1996, 2000) and others. There are $N = 171$ rice farms and $T = 6$ six-month growing seasons. Output is rice in kilograms and inputs are land in hectares, labor in hours, seed in kilograms and two types of fertilizer (urea in kilograms and phosphate in kilograms). The functional form is Cobb-Douglas with some dummy variables added for region, seasonality for dry or wet season, the use of pesticide and seed types for high yield or traditional or mixed. For a complete discussion of the data, see Erwidodo (1990).

The estimated regression parameters are given in Horrace and Schmidt (1996) and we will not repeat them here. Instead we will give point estimates of efficiencies (r_i^*) and 90% confidence intervals for these efficiencies. There are 171 firms and so we report results for the three firms (164, 118, and 163) that are most efficient; for the firms at the 75^{th} percentile (31), 50^{th} percentile (15) and 25^{th} percentile (16) of the efficiency distribution; and for the two worst firms (117, 45). All of these rankings are according to the fixed effects estimates.

We begin with Table 9. It gives the fixed effects point estimates and the lower and upper bounds of the 90% feasible parametric confidence intervals. For the purpose of comparison we also give the point estimates and the lower and upper bound of the 90% confidence intervals for the MLE based on the assumption that inefficiency has a half-normal distribution. See Horrace and Schmidt (1996) for the details of the calculations for the MLE.

The estimated efficiency levels based on the fixed effects estimates are rather low. They are certainly much smaller than the MLE estimates. This is presumably due to bias in the fixed effects estimates, as discussed previously. This data set has characteristics that should make the bias problem severe: N is large; the α_i are estimated imprecisely because σ_v^2 is large and T is small; and there are near ties for $\max_j \alpha_j$ because σ_u^2 is small.

Table 10 gives 90% confidence intervals based on the percentile bootstrap, the iterated bootstrap, and four methods that attempt to correct for bias (the BC_a bootstrap, the bias-corrected bootstrap, the bias-corrected bootstrap with BCa, and Hall's percentile method). All of these results are based on the indirect method. The bootstrap results are based on 1000 replications, and in the case of the iterated bootstrap each second-level bootstrap is also based on 1000 replications.

There is some similarity between the intervals from different methods, but there are also some interesting comparisons to make. The percentile bootstrap intervals are clearly closest to zero (i.e. they would indicate the lowest levels of efficiency). This is presumably a reflection of bias. Note, for example, that the midpoints of these intervals are less than the fixed effects estimate (which is itself biased toward zero). For the reasons given above, we do not regard these intervals as trustworthy for this data set. The iterated bootstrap intervals are centered similarly to the percentile bootstrap but are wider. They are about as wide as the feasible parametric intervals. The four methods that correct for bias give results that are relatively similar to each other. Compared to the feasible parametric method, the percentile bootstrap or the iterated percentile bootstrap, they clearly indicate higher efficiency levels. Presumably this is because they are less biased than those methods (which do not make a bias correction). However, given the results of our simulations, we still expect the bias corrected methods to be biased, for this data set. This is consistent with the fact that they show considerably lower efficiency levels than the intervals based on the MLE.

Table 11 gives the confidence intervals for the direct method. (We do not repeat the results for those methods for which the direct and indirect methods give the same results.) The direct and indirect methods give similar results, though the direct method yields intervals that indicate slightly lower efficiency levels.

The characteristics of this data set (large N, small T, small variance of inefficiency relative to noise) are such that there may not be a satisfactory alternative to making a distributional assumption on inefficiency and doing MLE. If one is not prepared to make a distributional assumption, a bias corrected version of the bootstrap would be recommended.

7.2 Texas Utilities

In this section, we consider the Texas utility data of Kumbhakar (1996), which was also analyzed by Horrace and Schmidt (1996, 2000). As in the previous section, we will estimate a Cobb-Douglas production function, whereas Kumbhakar (1996) estimated a cost function. The data contain information on output and inputs of 10 privately owned Texas electric utilities for 18 years from 1966 to 1983. Output is electric power generated, and input measures on labor, capital and fuel are derived from dividing expenditures on each input by its price. For more details on the data see Kumbhakar (1996).

Table 12 gives the fixed effects point estimates, the 90% parametric intervals, and the MLE point estimates and 90% confidence intervals. The format is the same as that of Table 9, except that now we can report the results for all of the firms. Tables 13 and 14 give the confidence intervals for the same set of procedures as before, and they are of the same format as Tables 10 and 11, except that results are given for all firms.

Compared to the previous data set, we estimate the intercepts α_i much more precisely, because T is larger and σ_v^2 is smaller. For this reason, and also because N is smaller, we expect there not to be a severe finite sample bias problem in the fixed effects estimates, and we expect that the choice of technique will not matter as much.

The MLE estimated efficiencies are larger than those based on fixed effects (except for the "best" firm), but the difference is not nearly as large as for the previous data set. Similarly, the MLE confidence intervals are narrower than the feasible parametric intervals, but not by nearly as much as in Table 9. A distributional assumption is much less valuable in the present case. In fact, the accuracy of the MLE intervals is now suspect, because we have only 10 firms, and the asymptotic justification for the MLE requires large N.

Comparing the results in Tables 12, 13, and 14, we can see that the parametric intervals and all of the bootstrapping intervals are quite similar. The bias problem is apparently negligible for this data set, and correspondingly our faith in the accuracy of these intervals is relatively strong.

We can compare the features of this data set with the setup of our simulation. One of the parametric configurations in our simulation had $T = 20$, $\gamma^* = 0.5$, and $N = 10$, which matches these data reasonably well. In that case the coverage rates of the various confidence intervals were in the range of 0.87 to 0.89, which are obviously close to 0.90, indicating that the confidence intervals are quite reliable.

8 Conclusions

In this paper we have provided a survey of the use of bootstrapping to construct confidence intervals for efficiency measures. We discussed several versions of the bootstrap, including the percentile bootstrap, the iterated bootstrap, the bias-adjusted and accelerated bootstrap, the bias-corrected bootstrap, and Hall's percentile method. In stochastic frontier models, these methods can be applied to the fixed effects estimates, yielding inferences that are correct asymptotically as $T \to \infty$ with N fixed.

We have proposed a simple parametric method of constructing confidence intervals. It uses standard methods and simply acts as if the identity of the best firm is known. This procedure is valid under the same conditions that the bootstrap methods are valid, namely, as $T \to \infty$ with N fixed, and provided that there is a unique best firm.

The main problem that we encounter is the upward bias in the fixed effects estimate of the frontier, which translates into a downward bias for the estimated efficiencies. The bias is large when T is small, N is large, and/or statistical noise is large relative to the variation in the frontier. These are exactly the same circumstances in which the identity of the best firm is uncertain, and so it is fair to say that bias is a problem when the identity of the best firm is in question.

Our simulation results show that the percentile bootstrap is seriously inaccurate when the bias problem exists. The percentile bootstrap intervals are mis-centered because the bias in the original estimates is compounded by similar "bias" in the bootstrap estimates. Our parametric intervals avoid the second source of bias and are more reliable than the percentile bootstrap intervals. However, when bias is a problem, they are still not very reliable. The BC_a bootstrap, bias-corrected bootstrap and Hall's percentile method all use the extent of the second-level bias in the percentile bootstrap (which is observable) to correct the bias in the fixed-effects estimates. When bias is a problem, these methods are a considerable improvement over the methods that do not correct for bias. However, a negative conclusion of the simulations is that their bias correction is only partially successful. It should be remembered that all of these methods are valid only for large T, and when T is not large enough that the identity of the best firm is clear, none of them will really be reliable. In such cases it may be worthwhile to consider assuming a distribution for technical inefficiency and using MLE.

We performed an empirical analysis of two data sets, one of which had characteristics very unfavorable to the bootstrap (large N, small T, and large variance of noise). In this case there was evidence of bias, and the bootstrap intervals were both unreliable and too wide to be informative. Our other data set had more favorable characteristics, and the empirical analysis yielded results that were quite precise and seemingly sensible. Hence, as in the simulations, a major lesson is that the reliability of bootstrap inference on efficiencies can be judged based on observable features of the data.

References

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.

Coelli, T.J., 1995, Estimators and hypothesis tests for a stochastic frontier function: A Monte Carlo analysis, *Journal of Productivity Analysis* 6, 247–268.

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

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

——— , 1985, Bootstrap confidence intervals for a class of parametric problems, *Biometrika* 72, 45–58.

——— , 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, Ph.D. dissertation, Department of Agricultural Economics, Michigan State University.

Hall, P., W. Härdle, and L. Simar, 1993, On the inconsistency of bootstrap distribution estimators, *Computational Statistics and Data Analysis* 16, 11–18.

——— , 1995, Iterated bootstrap with applications to frontier models, *Journal of Productivity Analysis* 6, 63–76.

Horrace, W.C., and P. Schmidt, 1996, Confidence statements for efficiency estimates from stochastic frontier models, *Journal of Productivity Analysis* 7, 257–282.

——— , 2000, Multiple comparisons with the best, with economic applications, *Journal of Applied Econometrics* 15, 1–26.

Kim, Y., 1999, A study in estimation and inference on firm efficiency, Ph.D. dissertation, Department of Economics, Michigan State University.

——— , and P. Schmidt, 1999, Marginal comparisons with the best and the efficiency measurement problem, Unpublished Manuscript, Department of Economics, Michigan State University.

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 (The Statistician)* 45, 319–335.

Lee, Y.H., 1991, Panel data models with multiplicative individual and time effects: Application to compensation and frontier production functions, Ph.D. dissertation, Department of Economics, Michigan State University.

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

Olson, J.A., P. Schmidt, and D. Waldman, 1980, A Monte Carlo study of estimators of stochastic frontier models, *Journal of Econometrics* 13, 67–82.

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.

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.

——— , and P. Wilson, 1998, Sensitivity of efficiency scores: How to bootstrap in nonparametric frontier models, *Management Science* 44, 49–61.

——— , 2000, A general methodology for bootstrapping in nonparametric frontier models, *Journal of Applied Statistics* 27, 779–802.

Table 1: Biases of Fixed Effects Estimates

T	γ^*	N	(1) $E(\hat{\alpha} - \alpha)$	(2) $E(\hat{\alpha} - \alpha_{(N)})$	(3) $E(\alpha_{(N)} - \alpha)$
10	0.1	10	0.102	0.131	-0.029
10	0.1	20	0.159	0.174	-0.015
10	0.1	50	0.230	0.236	-0.006
10	0.1	100	0.272	0.276	-0.004
10	0.5	10	-0.010	0.054	-0.064
10	0.5	20	0.047	0.082	-0.035
10	0.5	50	0.114	0.128	-0.014
10	0.5	100	0.151	0.158	-0.007
10	0.9	10	-0.075	0.010	-0.085
10	0.9	20	-0.029	0.017	-0.046
10	0.9	50	0.016	0.035	-0.019
10	0.9	100	0.039	0.049	-0.010
10	0.1	10	0.102	0.131	-0.029
20	0.1	10	0.046	0.076	-0.030
50	0.1	10	0.008	0.039	-0.031
100	0.1	10	-0.008	0.021	-0.029
10	0.5	10	-0.010	0.054	-0.064
20	0.5	10	-0.037	0.028	-0.065
50	0.5	10	-0.056	0.014	-0.070
100	0.5	10	-0.059	0.007	-0.066
10	0.9	10	-0.075	0.010	-0.085
20	0.9	10	-0.084	0.004	-0.088
50	0.9	10	-0.091	0.003	-0.094
100	0.9	10	-0.088	0.001	-0.089

Table 2: 90% Confidence Intervals for Relative Efficiency (r_i^*), Indirect Method

			Infeasible Parametric				Feasible Parametric				Percentile Bootstrap			
T	γ^*	N	width	$P_{(<lb)}$	$P_{(>ub)}$	cover	width	$P_{(<lb)}$	$P_{(>ub)}$	cover	width	$P_{(<lb)}$	$P_{(>ub)}$	cover
10	0.1	10	0.547	0.048	0.044	0.908	0.467	0.070	0.107	0.823	0.356	0.002	0.283	0.715
10	0.1	20	0.574	0.048	0.047	0.905	0.471	0.041	0.154	0.805	0.347	0.000	0.449	0.551
10	0.1	50	0.588	0.050	0.050	0.900	0.457	0.018	0.249	0.733	0.324	0.000	0.672	0.328
10	0.1	100	0.593	0.048	0.047	0.905	0.445	0.010	0.326	0.664	0.305	0.000	0.801	0.199
10	0.5	10	0.324	0.048	0.044	0.908	0.303	0.055	0.067	0.878	0.251	0.014	0.149	0.838
10	0.5	20	0.340	0.048	0.047	0.905	0.309	0.035	0.093	0.872	0.246	0.003	0.245	0.752
10	0.5	50	0.348	0.050	0.050	0.900	0.303	0.017	0.155	0.828	0.231	0.001	0.444	0.555
10	0.5	100	0.351	0.048	0.047	0.905	0.297	0.009	0.211	0.780	0.219	0.000	0.602	0.398
10	0.9	10	0.127	0.048	0.044	0.908	0.125	0.048	0.051	0.901	0.112	0.034	0.083	0.883
10	0.9	20	0.133	0.048	0.047	0.905	0.130	0.036	0.062	0.901	0.113	0.019	0.125	0.856
10	0.9	50	0.136	0.050	0.050	0.900	0.131	0.019	0.097	0.884	0.109	0.005	0.235	0.760
10	0.9	100	0.137	0.048	0.047	0.905	0.130	0.010	0.130	0.860	0.105	0.001	0.361	0.637
10	0.1	10	0.547	0.048	0.044	0.908	0.467	0.070	0.107	0.823	0.356	0.002	0.283	0.715
20	0.1	10	0.379	0.044	0.042	0.913	0.345	0.066	0.086	0.848	0.282	0.002	0.213	0.785
50	0.1	10	0.235	0.042	0.048	0.910	0.225	0.057	0.080	0.863	0.196	0.006	0.166	0.829
100	0.1	10	0.165	0.041	0.041	0.918	0.161	0.050	0.066	0.884	0.145	0.008	0.128	0.864
10	0.5	10	0.324	0.048	0.044	0.908	0.303	0.055	0.067	0.878	0.251	0.014	0.149	0.838
20	0.5	10	0.228	0.044	0.042	0.913	0.219	0.051	0.061	0.888	0.193	0.014	0.109	0.877
50	0.5	10	0.142	0.042	0.048	0.910	0.140	0.047	0.059	0.894	0.130	0.021	0.087	0.892
100	0.5	10	0.100	0.041	0.041	0.918	0.099	0.043	0.051	0.906	0.094	0.023	0.073	0.904
10	0.9	10	0.127	0.048	0.044	0.908	0.125	0.048	0.051	0.901	0.112	0.034	0.083	0.883
20	0.9	10	0.090	0.044	0.042	0.913	0.089	0.049	0.048	0.903	0.083	0.034	0.067	0.899
50	0.9	10	0.057	0.042	0.048	0.910	0.056	0.043	0.051	0.906	0.055	0.034	0.062	0.903
100	0.9	10	0.040	0.041	0.041	0.918	0.040	0.041	0.044	0.915	0.039	0.034	0.053	0.913

Table 3: 90% Confidence Intervals for Relative Efficiency (r_i^*), Indirect Method

			BC_a Bootstrap				Bias Corrected Bootstrap				Bias Corrected Bootstrap with BC_a				Hall's Percentile Method			
T	γ^*	N	width	$P_{(<lb)}$	$P_{(>ub)}$	cover	width	$P_{(<lb)}$	$P_{(>ub)}$	cover	width	$P_{(<lb)}$	$P_{(>ub)}$	cover	width	$P_{(<lb)}$	$P_{(>ub)}$	cover
10	0.1	10	0.337	0.015	0.118	0.867	0.414	0.028	0.112	0.860	0.432	0.013	0.117	0.870	0.430	0.117	0.104	0.778
10	0.1	20	0.333	0.014	0.156	0.830	0.418	0.020	0.142	0.838	0.424	0.012	0.148	0.840	0.427	0.073	0.133	0.794
10	0.1	50	0.327	0.009	0.234	0.757	0.406	0.011	0.205	0.784	0.406	0.008	0.212	0.780	0.411	0.035	0.193	0.772
10	0.1	100	0.314	0.007	0.306	0.687	0.395	0.007	0.247	0.746	0.394	0.006	0.254	0.740	0.399	0.020	0.233	0.746
10	0.5	10	0.255	0.039	0.086	0.874	0.271	0.061	0.078	0.861	0.286	0.037	0.086	0.877	0.277	0.110	0.073	0.817
10	0.5	20	0.247	0.037	0.106	0.857	0.273	0.045	0.098	0.857	0.279	0.035	0.103	0.862	0.278	0.086	0.092	0.822
10	0.5	50	0.237	0.025	0.158	0.817	0.265	0.027	0.145	0.828	0.266	0.023	0.152	0.825	0.268	0.048	0.135	0.816
10	0.5	100	0.228	0.018	0.200	0.782	0.258	0.018	0.178	0.804	0.258	0.016	0.185	0.799	0.260	0.031	0.167	0.802
10	0.9	10	0.116	0.055	0.086	0.859	0.114	0.090	0.060	0.849	0.120	0.055	0.085	0.861	0.114	0.093	0.059	0.848
10	0.9	20	0.114	0.058	0.082	0.860	0.116	0.072	0.073	0.855	0.120	0.055	0.081	0.864	0.117	0.090	0.070	0.840
10	0.9	50	0.109	0.050	0.110	0.840	0.113	0.052	0.102	0.846	0.115	0.047	0.107	0.846	0.114	0.069	0.096	0.835
10	0.9	100	0.106	0.035	0.136	0.829	0.111	0.035	0.127	0.839	0.111	0.032	0.132	0.836	0.111	0.047	0.118	0.835
10	0.1	10	0.337	0.015	0.118	0.867	0.414	0.028	0.112	0.860	0.432	0.013	0.117	0.870	0.430	0.117	0.104	0.778
20	0.1	10	0.268	0.027	0.087	0.886	0.313	0.040	0.083	0.877	0.327	0.024	0.087	0.890	0.321	0.126	0.077	0.797
50	0.1	10	0.190	0.034	0.085	0.882	0.207	0.050	0.076	0.874	0.218	0.032	0.081	0.887	0.210	0.114	0.071	0.815
100	0.1	10	0.144	0.034	0.072	0.894	0.150	0.053	0.062	0.885	0.158	0.032	0.071	0.897	0.151	0.101	0.059	0.840
10	0.5	10	0.255	0.039	0.086	0.874	0.271	0.061	0.078	0.861	0.286	0.037	0.086	0.877	0.277	0.110	0.073	0.817
20	0.5	10	0.197	0.045	0.076	0.879	0.202	0.070	0.063	0.867	0.215	0.045	0.075	0.880	0.205	0.111	0.060	0.829
50	0.5	10	0.134	0.049	0.078	0.872	0.133	0.076	0.059	0.864	0.141	0.047	0.079	0.874	0.134	0.092	0.057	0.852
100	0.5	10	0.096	0.046	0.077	0.877	0.095	0.080	0.051	0.869	0.100	0.045	0.076	0.879	0.095	0.080	0.050	0.870
10	0.9	10	0.116	0.055	0.086	0.859	0.114	0.090	0.060	0.849	0.120	0.055	0.085	0.861	0.114	0.093	0.059	0.848
20	0.9	10	0.086	0.058	0.076	0.865	0.084	0.091	0.051	0.858	0.089	0.057	0.076	0.867	0.085	0.089	0.051	0.860
50	0.9	10	0.056	0.048	0.077	0.875	0.055	0.077	0.053	0.870	0.057	0.047	0.077	0.876	0.055	0.066	0.051	0.882
100	0.9	10	0.040	0.045	0.066	0.889	0.039	0.070	0.046	0.884	0.040	0.043	0.067	0.890	0.039	0.060	0.045	0.895

Table 4: Biases in Bootstrap Intervals

T	γ^*	N	(1) $\alpha_{(N)} =$ $\max_j \alpha_j$	(2) $\hat{\alpha}_{[N]} =$ $\max_j \hat{\alpha}_j$	(3) $\max_j \hat{\alpha}_j^{(b)}$	(4) (2)-(1)	(5) (3)-(2)	(6) (4)-(5)
10	0.1	10	0.971	1.100	1.174	0.129	0.074	0.055
50	0.1	10	0.969	1.007	1.035	0.039	0.028	0.011
10	0.1	50	0.994	1.230	1.341	0.236	0.111	0.125
10	0.5	10	0.936	0.989	1.026	0.053	0.037	0.016
50	0.5	10	0.930	0.944	0.955	0.014	0.011	0.003
10	0.5	50	0.986	1.114	1.181	0.128	0.067	0.061
10	0.9	10	0.915	0.924	0.932	0.009	0.008	0.001
50	0.9	10	0.906	0.909	0.911	0.003	0.002	0.001
10	0.9	50	0.981	1.016	1.038	0.035	0.021	0.014

Table 5: 90% Confidence Intervals for Relative Efficiency (r_i^*), Direct Method

			BC_a Bootstrap				Bias Corrected Bootstrap				Bias Corrected Bootstrap with BC_a				Hall's Percentile Method			
T	γ^*	N	width	$P_{(<lb)}$	$P_{(>ub)}$	cover	width	$P_{(<lb)}$	$P_{(>ub)}$	cover	width	$P_{(<lb)}$	$P_{(>ub)}$	cover	width	$P_{(<lb)}$	$P_{(>ub)}$	cover
10	0.1	10	0.334	0.015	0.126	0.859	0.356	0.038	0.172	0.790	0.373	0.025	0.143	0.833	0.356	0.111	0.220	0.669
10	0.1	20	0.328	0.014	0.173	0.813	0.347	0.030	0.227	0.743	0.355	0.026	0.195	0.779	0.347	0.070	0.280	0.650
10	0.1	50	0.319	0.009	0.264	0.727	0.324	0.018	0.333	0.649	0.330	0.018	0.296	0.686	0.324	0.034	0.393	0.573
10	0.1	100	0.306	0.007	0.339	0.654	0.305	0.014	0.407	0.579	0.310	0.015	0.369	0.616	0.305	0.019	0.466	0.514
10	0.5	10	0.253	0.038	0.090	0.872	0.251	0.065	0.106	0.830	0.267	0.048	0.094	0.858	0.251	0.098	0.138	0.764
10	0.5	20	0.243	0.037	0.113	0.850	0.246	0.052	0.138	0.810	0.253	0.049	0.120	0.831	0.246	0.077	0.173	0.751
10	0.5	50	0.231	0.025	0.174	0.800	0.231	0.035	0.209	0.756	0.234	0.036	0.189	0.775	0.231	0.043	0.247	0.709
10	0.5	100	0.220	0.017	0.225	0.758	0.219	0.024	0.263	0.713	0.221	0.027	0.240	0.733	0.219	0.028	0.303	0.669
10	0.9	10	0.115	0.054	0.088	0.858	0.112	0.087	0.069	0.844	0.119	0.057	0.077	0.865	0.112	0.082	0.087	0.831
10	0.9	20	0.113	0.058	0.085	0.857	0.113	0.073	0.084	0.844	0.117	0.061	0.084	0.856	0.113	0.079	0.096	0.825
10	0.9	50	0.108	0.050	0.116	0.834	0.109	0.055	0.121	0.824	0.110	0.054	0.117	0.829	0.109	0.063	0.132	0.805
10	0.9	100	0.103	0.035	0.147	0.818	0.105	0.038	0.151	0.811	0.105	0.038	0.147	0.815	0.105	0.043	0.161	0.795
10	0.1	10	0.334	0.015	0.126	0.859	0.356	0.038	0.172	0.790	0.373	0.025	0.143	0.833	0.356	0.111	0.220	0.669
20	0.1	10	0.265	0.026	0.093	0.881	0.282	0.047	0.116	0.837	0.296	0.035	0.100	0.865	0.282	0.119	0.147	0.734
50	0.1	10	0.189	0.034	0.088	0.879	0.196	0.054	0.094	0.852	0.206	0.039	0.089	0.873	0.196	0.108	0.111	0.781
100	0.1	10	0.144	0.033	0.075	0.892	0.145	0.055	0.074	0.871	0.153	0.035	0.074	0.891	0.145	0.096	0.087	0.817
10	0.5	10	0.253	0.038	0.090	0.872	0.251	0.065	0.106	0.830	0.267	0.048	0.094	0.858	0.251	0.098	0.138	0.764
20	0.5	10	0.196	0.045	0.078	0.877	0.193	0.072	0.079	0.849	0.206	0.052	0.078	0.870	0.193	0.097	0.101	0.802
50	0.5	10	0.134	0.049	0.081	0.870	0.130	0.076	0.069	0.856	0.138	0.051	0.077	0.872	0.130	0.083	0.085	0.832
100	0.5	10	0.096	0.046	0.079	0.875	0.094	0.078	0.058	0.864	0.099	0.047	0.069	0.884	0.094	0.071	0.071	0.858
10	0.9	10	0.115	0.054	0.088	0.858	0.112	0.087	0.069	0.844	0.119	0.057	0.077	0.865	0.112	0.082	0.087	0.831
20	0.9	10	0.086	0.058	0.079	0.862	0.083	0.089	0.057	0.854	0.088	0.059	0.069	0.873	0.083	0.079	0.070	0.851
50	0.9	10	0.056	0.048	0.079	0.873	0.055	0.075	0.056	0.869	0.057	0.048	0.069	0.884	0.055	0.060	0.067	0.873
100	0.9	10	0.039	0.044	0.069	0.887	0.039	0.068	0.048	0.884	0.040	0.043	0.058	0.899	0.039	0.055	0.057	0.888

Table 6: Biases of Fixed Effects Estimates (Case that u_i are fixed over replications)

T	γ^*	N	(1) $\mathrm{E}(\hat{\alpha}-\alpha)$	(2) $\mathrm{E}(\hat{\alpha}-\alpha_{(N)})$	(3) $\mathrm{E}(\alpha_{(N)}-\alpha)$
10	0.1	10	0.008	0.025	-0.017
10	0.1	20	0.021	0.037	-0.016
10	0.1	50	0.042	0.059	-0.017
10	0.1	100	0.065	0.082	-0.017
10	0.5	10	-0.036	0.001	-0.037
10	0.5	20	-0.035	0.002	-0.037
10	0.5	50	-0.037	0.000	-0.037
10	0.5	100	-0.036	0.001	-0.037
10	0.9	10	-0.049	0.000	-0.049
10	0.9	20	-0.049	0.001	-0.050
10	0.9	50	-0.049	0.000	-0.049
10	0.9	100	-0.049	0.000	-0.049
10	0.1	10	0.008	0.025	-0.017
20	0.1	10	-0.009	0.008	-0.017
50	0.1	10	-0.018	-0.002	-0.016
100	0.1	10	-0.017	-0.001	-0.016
10	0.5	10	-0.036	0.001	-0.037
20	0.5	10	-0.034	0.003	-0.037
50	0.5	10	-0.038	-0.001	-0.037
100	0.5	10	-0.037	-0.001	-0.036
10	0.9	10	-0.049	0.000	-0.049
20	0.9	10	-0.048	0.001	-0.049
50	0.9	10	-0.050	-0.001	-0.049
100	0.9	10	-0.050	0.000	-0.050

Table 7: 90% Confidence Intervals for Relative Efficiency (r_i^*), Indirect Method (Case that u_i^* are fixed across replications)

			Infeasible Parametric				Feasible Parametric				Percentile Bootstrap			
T	γ^*	N	width	$P_{(<lb)}$	$P_{(>ub)}$	cover	width	$P_{(<lb)}$	$P_{(>ub)}$	cover	width	$P_{(<lb)}$	$P_{(>ub)}$	cover
10	0.1	10	0.456	0.045	0.045	0.909	0.441	0.031	0.047	0.922	0.356	0.019	0.093	0.888
10	0.1	20	0.479	0.043	0.044	0.914	0.459	0.022	0.048	0.931	0.354	0.009	0.129	0.861
10	0.1	50	0.494	0.048	0.051	0.901	0.462	0.013	0.059	0.928	0.335	0.004	0.218	0.778
10	0.1	100	0.499	0.053	0.051	0.896	0.456	0.007	0.066	0.927	0.319	0.001	0.308	0.691
10	0.5	10	0.212	0.045	0.045	0.909	0.212	0.045	0.045	0.909	0.199	0.057	0.056	0.887
10	0.5	20	0.223	0.043	0.044	0.914	0.223	0.043	0.044	0.914	0.210	0.053	0.054	0.893
10	0.5	50	0.230	0.048	0.051	0.901	0.230	0.048	0.051	0.901	0.217	0.058	0.060	0.883
10	0.5	100	0.233	0.053	0.051	0.896	0.233	0.053	0.051	0.896	0.218	0.061	0.060	0.878
10	0.9	10	0.071	0.045	0.045	0.909	0.071	0.045	0.045	0.909	0.066	0.057	0.056	0.887
10	0.9	20	0.074	0.043	0.044	0.914	0.074	0.043	0.044	0.914	0.070	0.053	0.054	0.893
10	0.9	50	0.077	0.048	0.051	0.901	0.077	0.048	0.051	0.901	0.073	0.058	0.060	0.883
10	0.9	100	0.077	0.053	0.051	0.896	0.077	0.053	0.051	0.896	0.074	0.062	0.060	0.878
10	0.1	10	0.456	0.045	0.045	0.909	0.441	0.031	0.047	0.922	0.356	0.019	0.093	0.888
20	0.1	10	0.313	0.038	0.044	0.918	0.311	0.031	0.044	0.924	0.281	0.028	0.057	0.915
50	0.1	10	0.196	0.055	0.049	0.896	0.196	0.054	0.049	0.896	0.191	0.053	0.051	0.896
100	0.1	10	0.138	0.045	0.041	0.914	0.138	0.045	0.041	0.914	0.137	0.047	0.041	0.912
10	0.5	10	0.212	0.045	0.045	0.909	0.212	0.045	0.045	0.909	0.199	0.057	0.056	0.887
20	0.5	10	0.147	0.038	0.044	0.918	0.147	0.038	0.044	0.918	0.143	0.043	0.048	0.909
50	0.5	10	0.093	0.055	0.049	0.896	0.093	0.055	0.049	0.896	0.092	0.056	0.051	0.894
100	0.5	10	0.065	0.045	0.041	0.914	0.065	0.045	0.041	0.914	0.065	0.047	0.041	0.912
10	0.9	10	0.071	0.045	0.045	0.909	0.071	0.045	0.045	0.909	0.066	0.057	0.056	0.887
20	0.9	10	0.050	0.038	0.044	0.918	0.050	0.038	0.044	0.918	0.048	0.043	0.048	0.909
50	0.9	10	0.031	0.055	0.049	0.896	0.031	0.055	0.049	0.896	0.031	0.056	0.051	0.894
100	0.9	10	0.022	0.045	0.041	0.914	0.022	0.045	0.041	0.914	0.022	0.047	0.041	0.912

Table 8: 90% Confidence Intervals for Relative Efficiency (r_i^*), Indirect Method (Case that u_i^* are fixed across replications)

			BC_a Bootstrap				Bias Corrected bootstrap				Bias Corrected Bootstrap with BC_a				Hall's Percentile Method			
T	γ^*	N	width	$P_{(<lb)}$	$P_{(>ub)}$	cover	width	$P_{(<lb)}$	$P_{(>ub)}$	cover	width	$P_{(<lb)}$	$P_{(>ub)}$	cover	width	$P_{(<lb)}$	$P_{(>ub)}$	cover
10	0.1	10	0.355	0.099	0.068	0.833	0.396	0.123	0.058	0.819	0.414	0.093	0.067	0.840	0.408	0.146	0.055	0.799
10	0.1	20	0.350	0.113	0.062	0.825	0.409	0.121	0.057	0.822	0.416	0.105	0.061	0.834	0.418	0.137	0.053	0.810
10	0.1	50	0.339	0.127	0.069	0.805	0.410	0.126	0.062	0.812	0.410	0.118	0.066	0.817	0.415	0.134	0.057	0.809
10	0.1	100	0.330	0.116	0.075	0.809	0.403	0.112	0.064	0.823	0.402	0.108	0.068	0.824	0.408	0.119	0.058	0.822
10	0.5	10	0.200	0.060	0.086	0.854	0.199	0.089	0.055	0.856	0.200	0.060	0.086	0.855	0.199	0.058	0.056	0.886
10	0.5	20	0.211	0.056	0.073	0.871	0.210	0.074	0.054	0.872	0.211	0.056	0.073	0.871	0.211	0.056	0.054	0.891
10	0.5	50	0.217	0.061	0.071	0.868	0.217	0.072	0.060	0.868	0.217	0.061	0.071	0.868	0.217	0.062	0.059	0.878
10	0.5	100	0.219	0.068	0.066	0.866	0.219	0.074	0.061	0.866	0.219	0.068	0.066	0.866	0.219	0.071	0.061	0.869
10	0.9	10	0.066	0.060	0.056	0.884	0.066	0.058	0.055	0.887	0.066	0.059	0.055	0.885	0.066	0.057	0.056	0.887
10	0.9	20	0.070	0.054	0.053	0.892	0.070	0.053	0.053	0.894	0.070	0.055	0.053	0.892	0.070	0.053	0.054	0.893
10	0.9	50	0.073	0.059	0.060	0.881	0.073	0.058	0.060	0.883	0.073	0.059	0.059	0.881	0.073	0.058	0.059	0.883
10	0.9	100	0.073	0.063	0.061	0.876	0.074	0.062	0.060	0.878	0.073	0.063	0.060	0.876	0.074	0.062	0.061	0.877
10	0.1	10	0.355	0.099	0.068	0.833	0.396	0.123	0.058	0.819	0.414	0.093	0.067	0.840	0.408	0.146	0.055	0.799
20	0.1	10	0.291	0.080	0.076	0.845	0.292	0.113	0.049	0.838	0.307	0.075	0.076	0.849	0.296	0.103	0.048	0.849
50	0.1	10	0.196	0.066	0.110	0.824	0.192	0.128	0.051	0.821	0.197	0.066	0.111	0.824	0.192	0.070	0.051	0.879
100	0.1	10	0.138	0.048	0.068	0.884	0.137	0.074	0.040	0.886	0.138	0.048	0.068	0.883	0.138	0.046	0.041	0.913
10	0.5	10	0.200	0.060	0.086	0.854	0.199	0.089	0.055	0.856	0.200	0.060	0.086	0.855	0.199	0.058	0.056	0.886
20	0.5	10	0.143	0.044	0.049	0.907	0.143	0.045	0.047	0.907	0.143	0.044	0.049	0.907	0.143	0.044	0.048	0.908
50	0.5	10	0.092	0.058	0.052	0.890	0.092	0.057	0.051	0.892	0.092	0.058	0.052	0.890	0.092	0.056	0.051	0.894
100	0.5	10	0.065	0.048	0.040	0.912	0.065	0.046	0.040	0.914	0.065	0.048	0.041	0.911	0.065	0.045	0.041	0.913
10	0.9	10	0.066	0.060	0.056	0.884	0.066	0.058	0.055	0.887	0.066	0.059	0.055	0.885	0.066	0.057	0.056	0.887
20	0.9	10	0.048	0.044	0.047	0.908	0.048	0.044	0.047	0.909	0.048	0.044	0.048	0.908	0.048	0.044	0.048	0.908
50	0.9	10	0.031	0.058	0.052	0.890	0.031	0.057	0.051	0.892	0.031	0.058	0.052	0.890	0.031	0.056	0.051	0.894
100	0.9	10	0.022	0.048	0.040	0.912	0.022	0.046	0.040	0.914	0.022	0.048	0.041	0.911	0.022	0.045	0.041	0.913

Table 9: Estimated Efficiencies and 90% Confidence Intervals: Indonesian Rice Farms

Firm No.	Fixed Effects Point Estimate	Fixed Effects LB	Fixed Effects UB	MLE Point Estimate	MLE LB	MLE UB
164	1.000	1.000	1.000	0.964	0.903	0.998
118	0.933	0.682	1.000	0.964	0.902	0.998
163	0.932	0.682	1.000	0.959	0.890	0.997
⋮	⋮	⋮	⋮	⋮	⋮	⋮
31	0.620	0.447	0.859	0.924	0.823	0.994
⋮	⋮	⋮	⋮	⋮	⋮	⋮
15	0.554	0.403	0.762	0.923	0.792	0.990
⋮	⋮	⋮	⋮	⋮	⋮	⋮
16	0.501	0.362	0.694	0.845	0.725	0.969
⋮	⋮	⋮	⋮	⋮	⋮	⋮
117	0.380	0.275	0.524	0.773	0.658	0.907
45	0.366	0.266	0.504	0.774	0.659	0.908

Table 10: 90% Confidence Intervals by Indirect Method: Indonesian Rice Farms

Firm No.	FE Est.	Percentile Bootstrap		Iterated Bootstrap		BC_a Bootstrap		Bias Corrected Bootstrap		Bias Corrected with BC_a		Hall's Percentile Method	
		LB	UB	LB	UB	LB	UB	LB	UB	LB	UB	LB	UB
164	1.000	0.743	1.000	0.663	1.000	0.807	1.000	0.876	1.000	0.788	1.000	1.000	1.000
118	0.933	0.672	1.000	0.582	1.000	0.775	1.000	0.796	1.000	0.770	1.000	0.871	1.000
163	0.932	0.683	1.000	0.600	1.000	0.776	1.000	0.801	1.000	0.770	1.000	0.868	1.000
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
31	0.620	0.446	0.750	0.396	0.805	0.512	0.824	0.520	0.875	0.517	0.871	0.513	0.862
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
15	0.554	0.400	0.638	0.362	0.692	0.477	0.720	0.469	0.749	0.464	0.740	0.482	0.768
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
16	0.501	0.358	0.582	0.320	0.636	0.421	0.649	0.423	0.687	0.411	0.678	0.431	0.700
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
117	0.380	0.274	0.446	0.252	0.498	0.320	0.509	0.318	0.519	0.318	0.518	0.323	0.527
45	0.366	0.267	0.424	0.237	0.465	0.309	0.508	0.313	0.498	0.313	0.497	0.316	0.502

Table 11: 90% Confidence Intervals by Direct Method: Indonesian Rice Farms

Firm No.	FE Est.	Bias Corrected Bootstrap LB	UB	Bias Corrected with BC_a LB	UB	Hall's Percentile Method LB	UB
164	1.000	0.892	1.000	0.825	1.000	1.000	1.000
118	0.933	0.809	1.000	0.796	1.000	0.867	1.000
163	0.932	0.812	1.000	0.801	1.000	0.864	1.000
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
31	0.620	0.524	0.828	0.531	0.840	0.490	0.794
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
15	0.554	0.475	0.713	0.476	0.717	0.471	0.709
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
16	0.501	0.428	0.652	0.427	0.651	0.419	0.644
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
117	0.380	0.321	0.494	0.327	0.503	0.314	0.486
45	0.366	0.316	0.473	0.318	0.478	0.308	0.465

Table 12: Estimated Efficiencies and 90% Confidence Intervals: Texas Utilities

Firm No.	Fixed Effects Point Estimate	LB	UB	MLE Point Estimate	LB	UB
5	1.000	1.000	1.000	0.987	0.971	0.999
3	0.916	0.823	1.000	0.978	0.959	0.996
10	0.861	0.786	0.943	0.908	0.889	0.927
1	0.835	0.784	0.889	0.864	0.846	0.882
8	0.820	0.773	0.869	0.846	0.828	0.864
9	0.806	0.766	0.848	0.826	0.809	0.843
2	0.801	0.749	0.855	0.831	0.814	0.848
7	0.786	0.732	0.844	0.817	0.800	0.834
6	0.785	0.730	0.845	0.820	0.803	0.837
4	0.762	0.719	0.808	0.786	0.770	0.801

Table 13: 90% Confidence Intervals by Indirect Method: Texas Utilities

Firm No.	FE Est.	Percentile Bootstrap		Iterated Bootstrap		BC_a Bootstrap		Bias Corrected Bootstrap		Bias Corrected with BC_a		Hall's Percentile Method	
		LB	UB	LB	UB	LB	UB	LB	UB	LB	UB	LB	UB
5	1.000	0.988	1.000	0.997	1.000	0.907	1.000	0.992	1.000	0.911	1.000	1.000	1.000
3	0.916	0.827	1.000	0.835	1.000	0.825	1.000	0.828	1.000	0.823	1.000	0.840	1.000
10	0.861	0.793	0.924	0.797	0.920	0.790	0.922	0.794	0.924	0.789	0.923	0.802	0.934
1	0.835	0.788	0.877	0.792	0.874	0.783	0.874	0.789	0.878	0.784	0.875	0.794	0.884
8	0.820	0.777	0.859	0.781	0.857	0.777	0.859	0.778	0.860	0.777	0.860	0.782	0.865
9	0.806	0.769	0.841	0.775	0.838	0.773	0.843	0.771	0.842	0.772	0.843	0.773	0.845
2	0.801	0.753	0.842	0.758	0.839	0.754	0.842	0.755	0.844	0.753	0.843	0.761	0.851
7	0.786	0.736	0.830	0.739	0.827	0.736	0.830	0.738	0.833	0.736	0.831	0.744	0.839
6	0.785	0.732	0.832	0.737	0.828	0.731	0.831	0.734	0.835	0.731	0.831	0.741	0.843
4	0.762	0.721	0.798	0.726	0.795	0.718	0.797	0.722	0.800	0.719	0.797	0.727	0.806

Table 14: 90% Confidence Intervals by Direct Method: Texas Utilities

Firm No.	FE Est.	Bias Corrected Bootstrap LB	Bias Corrected Bootstrap UB	Bias Corrected with BC_a LB	Bias Corrected with BC_a UB	Hall's Percentile Method LB	Hall's Percentile Method UB
5	1.000	0.992	1.000	0.911	1.000	1.000	1.000
3	0.916	0.825	0.998	0.823	0.998	0.833	1.000
10	0.861	0.791	0.922	0.790	0.921	0.798	0.928
1	0.835	0.788	0.877	0.783	0.874	0.792	0.881
8	0.820	0.777	0.859	0.777	0.859	0.780	0.863
9	0.806	0.770	0.842	0.772	0.843	0.772	0.843
2	0.801	0.754	0.843	0.753	0.843	0.759	0.848
7	0.786	0.737	0.831	0.736	0.831	0.741	0.836
6	0.785	0.733	0.832	0.731	0.831	0.739	0.839
4	0.762	0.721	0.799	0.718	0.797	0.726	0.804