ESTIMATION AND INFERENCE IN PARAMETRIC

DETERMINISTIC FRONTIER MODELS

Christine Amsler
Michigan State University

Michael Leonard
Michigan State University

Peter Schmidt
Michigan State University
Yonsei University

April 3, 2011

ABSTRACT

In this paper we consider parametric deterministic frontier models. For example, the production frontier may be linear in the inputs, and the error is purely one-sided, with a known distribution such as exponential or half-normal.

The literature contains many negative results for this model. Schmidt (1976) showed that the Aigner-Chu (1968) linear programming estimator was the exponential MLE, but that this was a non-regular problem in which the statistical properties of the MLE were uncertain. Richmond (1974) and Greene (1980) showed how the model could be estimated by two different versions of corrected OLS, but this did not lead to methods of inference for the inefficiencies. Greene (1980) considered conditions on the distribution of inefficiency that make this a regular estimation problem, but many distributions that would be assumed do not satisfy these conditions.

In this paper we show that exact (finite sample) inference is possible when the frontier and the distribution of the one-sided error are known up to the values of some parameters. We give a number of analytical results for the case of intercept only with exponential errors. In other cases that include regressors or error distributions other than exponential, exact inference is still possible but simulation is needed to calculate the critical values.

We also discuss the case that the distribution of the error is unknown. In this case asymptotically valid inference is possible using resampling (subsampling or bootstrap) methods.

1. INTRODUCTION

In this paper we consider a linear deterministic frontier model of the form:

(1) $$y_i = \alpha + x_i'\beta - u_i \quad , \quad u_i \geq 0 \quad , \quad i = 1, 2, \ldots, n \quad .$$

We are mostly interested in the case in which the u_i are iid with a distribution that is known up to the value of some parameters. For example, we will discuss in detail the case that the distribution of the u_i is exponential with mean μ. We have $y_i \leq \alpha + x_i'\beta$, which is the deterministic frontier, and the model is parametric in the sense that a parametric assumption is made both about the functional form of the frontier and the distribution of u_i. The ultimate point of models of this type is to obtain an estimate of u_i (for each *i*), which is interpreted as a measure of the technical inefficiency of firm *i*. For an expository treatment of parametric deterministic frontier models, see Kumbhakar and Lovell (2000), section 3.2.1

The literature contains many negative results for this model. Aigner and Chu (1968) suggested a linear programming estimator that minimized the sum of absolute errors in (1) subject to the constraint that $y_i \leq \alpha + x_i'\beta$ for all *i*. They also suggested a quadratic programming estimator that minimized the sum of squared errors subject to the same constraints. They did not establish any statistical properties for these estimators. Schmidt (1976) showed that the Aigner-Chu linear programming estimator was the MLE under the assumption that the errors u_i are exponential and that their quadratic programming estimator was the MLE under the assumption that the errors are half-normal. However, he noted that this was a non-regular problem because the range of the random variable y_i depends on the parameters, which violates one of the usual regularity conditions for MLE. Therefore the statistical properties of the MLE were uncertain. Greene (1980) gave a condition on the distribution of inefficiency that makes this a regular estimation problem, namely that the density of *u* at the point *u* = 0 should equal zero. Some

possible distributions for u (e.g. lognormal, or gamma with parameter ≥ 2) satisfy this condition, but many other distributions that might commonly be assumed (e.g., exponential or half-normal) do not. Another possible method of estimation is corrected ordinary least squares (COLS), in which β is estimated by ordinary least squares (OLS) applied to equation (1), and then α is estimated by shifting the line upward (keeping the slope the same) until it lies on one point and above the others. This was first suggested by Winsten (1957) – though it was literally a one-sentence suggestion – and then further developed by Greene (1980), who proved the consistency of the COLS estimators of α and β. However, while this enabled consistent estimation of the u_i, methods of inference (such as construction of confidence intervals) for the u_i are still not established thirty years later.

The lack of more positive results for this problem is surprising because asymptotically valid methods of inference are available for models that make weaker assumptions. For example, in the stochastic frontier model of Aigner, Lovell and Schmidt (1977) and Meeusen and van den Broeck (1977), which adds normal noise to the model (1) and therefore clearly makes the estimation problem more difficult, we can estimate the u_i and construct asymptotically valid confidence intervals for them. See Jondrow, Lovell, Materov and Schmidt (1982), Battese and Coelli (1988) and Horrace and Schmidt (1996) for an analytical approach. Alternatively, Simar and Wilson (2010) show that bootstrapping also provides asymptotically valid confidence intervals for u_i in this model.

Asymptotically valid confidence intervals can also be constructed for non-parametric deterministic frontier models in which no distributional assumption is made for the u_i and no functional form is assumed for the frontier function. Examples are the FDH (free disposal hull) model for which only monotonicity in x is assumed for the frontier function and the DEA (data

envelopment analysis) model in which monotonicity and concavity are assumed. Consistency and rates of convergence of the efficiency estimators were established by Kneip, Park and Simar (1998) for DEA and by Park, Simar and Wiener (2000) for FDH. The asymptotic distribution of the efficiency estimates was obtained by Park, Simar and Wiener (2000) for FDH and by Kneip, Simar and Wilson (2008) and Park, Jeong and Simar (2010) for DEA. These asymptotic distributions can be used to create asymptotically valid confidence intervals for the efficiency estimates. Asymptotically valid confidence intervals can also be achieved by bootstrapping. The existence of the boundary condition on y implies that the naive bootstrap is not consistent, but asymptotically valid bootstrapping procedures are given by Simar and Wilson (1998, 2000A, 2000B) and Kneip, Simar and Wilson (2008). Also other subsampling methods (notably the "m out of n bootstrap") have been proposed by Simar and Wilson (2009).

In this paper we ask the obvious question of whether we can do better in the case of a parametric deterministic frontier model, which intuitively should be possible because we are making stronger assumptions. The answer is yes. Under a "scaling condition" that holds for many commonly-assumed distributions, including half-normal and exponential, we can construct exact confidence intervals in finite samples. This is not possible in stochastic frontier models or in non-parametric models like DEA.

The plan of the paper is as follows. In Section 2 we consider a very simple case in which there are no regressors (only intercept) and the error is exponential. This case allows analytical expressions for bias-corrected estimates and for confidence intervals that are valid in finite samples. In Section 3 we allow distributions other than exponential. So long as the distribution of u satisfies the "scaling property" that u is distributed as a scalar multiple of a random variable with a known density, we can still construct confidence intervals that are valid in finite samples.

However, now the critical values that are needed for the confidence intervals must be calculated by simulation. In Section 4, we add regressors to the model. Once again we can construct confidence intervals that are valid in finite samples, but now the critical values depend on the values of the regressors and must be calculated (observation by observation) by simulation. Section 5 considers the case that the distribution of *u* is unknown. Now we cannot construct confidence intervals that are valid in finite samples, but asymptotically valid inference is generally possible using resampling techniques (subsampling or the bootstrap). Finally, Section 6 contains our concluding remarks.

2. INTERCEPT ONLY, EXPONENTIAL ERRORS

In this section we consider the case of an intercept only (no regressors) and exponential errors. The point of considering this simple and empirically uninteresting case is that we can obtain analytical results that we cannot obtain in more complicated models. Thus we consider the model

(2) $$y_i = \alpha - u_i \quad , \quad u_i \geq 0 \quad , \quad i = 1, 2, \ldots, n \quad ,$$

where the u_i are iid as exponential with mean μ. The density of *u* is

(3) $$f(u) = \left(\frac{1}{\mu}\right) \exp\left(-\frac{u}{\mu}\right) .$$

This model is closely related to the "two parameter exponential distribution," which would correspond in our notation to $y_i = \alpha + u_i$, $\alpha > 0$, u_i exponential with mean μ. A textbook treatment of that model is given by, e.g., Lawless (1982), section 3.5. The results from that model apply to our model with a few sign changes.

Define $y_{max} = max_{i=1,\ldots,n}\, y_i$ and $u_{min} = min_{i=1,\ldots,n}\, u_i$, and note that $y_{max} = \alpha - u_{min}$. Also let $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ and similarly for $\bar{u}$. Then we have the following (well known) result.

RESULT 1. The MLE's of α and μ are $\hat{\alpha} = y_{max}$ and $\hat{\mu} = \hat{\alpha} - \bar{y} = y_{max} - \bar{y} = \bar{u} - u_{min}$.

Proofs not given in the text can be found in the Appendix.

It is clear that $\hat{\alpha}$ is biased downward, because $\hat{\alpha} = y_{max} < \alpha$ with probability one. This implies that $\hat{\mu}$ is also biased downward. However, these biases are easily evaluated, so that we can construct bias-adjusted (i.e. unbiased) estimators, as follows.

RESULT 2. The following estimators are unbiased for α and μ:

$$\tilde{\mu} = \frac{n}{n-1}(y_{max} - \bar{y}) = \frac{n}{n-1}\hat{\mu}$$

$$\tilde{\alpha} = \frac{n}{n-1}y_{max} - \frac{1}{n-1}\bar{y} = \hat{\alpha} + \frac{1}{n-1}\hat{\mu}$$

We now turn to the question of inference on the u_i. Since $u_i = \alpha - y_i$, it is natural to consider estimating it by $\hat{u}_i = \hat{\alpha} - y_i$ or $\tilde{u}_i = \tilde{\alpha} - y_i$. The second of these is unbiased in the sense that $E(\tilde{u}_i - u_i) = 0$. However, for purposes of construction of confidence intervals, it is simpler to consider the biased estimator $\hat{u}_i$, and (as we will point out later) we would get the same confidence intervals either way. We note that

(4) $$\hat{u}_i - u_i = \hat{\alpha} - \alpha = -u_{min}$$

so that the estimation error $\hat{u}_i - u_i$ is the same for all *i*. In fact, from (4) it is clear that the problem of construction of a confidence interval for u_i (based on $\hat{u}_i$) is exactly the same as the problem of construction of a confidence interval for α (based on $\hat{\alpha}$).

Since u_{min} is the minimum of a random sample of *n* exponential random variables with mean μ, it is distributed as exponential with mean $\frac{\mu}{n}$. This leads to the probability statement

(5) $$P(0.0253\frac{\mu}{n} \leq u_{min} \leq 3.689\frac{\mu}{n}) = 0.95\ ,$$

where $0.0253 = -ln(1 - 0.025)$ and $3.689 = -ln(1 - 0.975)$ are the appropriate (0.025 and 0.975) quantiles of the standard ($\mu = 1$) exponential distribution. Substituting $u_{min} = u_i - \hat{u}_i$, we obtain the confidence interval:

(6) $$P(\hat{u}_i + 0.0253\frac{\mu}{n} \leq u_i \leq \hat{u}_i + 3.689\frac{\mu}{n}) = 0.95.$$

Several things about this result are worth pointing out. First, it holds for any individual *i*, but it also holds simultaneously for all *i*, because $u_i - \hat{u}_i = u_{min}$ for all *i*, that is, the estimation error in $\hat{u}_i$ is the same for every observation. Second, it implicitly recognizes the downward bias of $\hat{u}_i$ as an estimate of u_i; the confidence interval is not symmetric around $\hat{u}_i$. In fact the entire confidence interval lies to the right of $\hat{u}_i$. Third, this interval is not feasible in practice because it depends on μ, but it would remain valid asymptotically if a consistent estimate $\hat{\mu}$ (e.g. the MLE given in Result 1) is used in place of μ.

In order to construct a confidence interval that is valid in finite samples, we need to find a quantity that is *pivotal*, that is, whose distribution does not depend on the parameters α and μ. This is not hard to do, as the following result demonstrates.

RESULT 3. Define $S_n = \frac{u_i - \hat{u}_i}{\hat{\mu}}$. Then the distribution of S_n does not depend on α or μ.

We note that S_n is defined in terms of $\hat{u}_i$ and u_i , but in fact it does not depend on *i* because $u_i - \hat{u}_i = u_{min}$ for all *i*. It just depends on *n*.

A sidelight is that, if we had chosen to construct confidence intervals based on $\tilde{u}_i$ instead of $\hat{u}_i$, we would have been led to the pivotal quantity $\tilde{S}_n = \frac{u_i - \tilde{u}_i}{\tilde{\mu}}$. However, using the relationships in Result 2, it is easy to show that $\tilde{S}_n = \frac{n-1}{n} S_n - \frac{1}{n}$. So exactly the same confidence intervals would result if we use $\tilde{u}_i$ or $\hat{u}_i$.

Because S_n is pivotal, we should be able to find critical values a_n and b_n (which depend on *n* but not on α or μ) such that

(7) $$P(a_n \leq S_n \leq b_n) = 0.95 .$$

If so, then simple arithmetic yields the confidence interval

(8) $$P(\hat{u}_i + a_n\hat{\mu} \leq u_i \leq \hat{u}_i + b_n\hat{\mu}) = 0.95 .$$

As above, it is noteworthy that this confidence interval would hold individually for any *i*, but it also would hold simultaneously for all *i*, because the estimation error in $\hat{u}_i$ is the same for every observation.

We could calculate a_n and b_n by simulation. Because S_n is pivotal, we can arbitrarily set $\alpha = 0$ and $\mu = 1$ (or pick any other values), generate data $y_i = -u_i$, *i* = 1,…,*n*, where the u_i are draws from a standard ($\mu = 1$) exponential distribution, and calculate S_n. (Equation (A3) of the Appendix gives a useful formula for calculating S_n.) Repeating this process a large number of times will reveal a_n and b_n as the 0.025 and 0.975 quantiles of the values of S_n. This is essentially a method of calculation, as opposed to estimation, because the results can be made as accurate as one wishes by using enough replications of the simulation.

However, in the present case, simulation is not actually needed because we can calculate the distribution of S_n analytically. The following result gives the asymptotic and finite sample distributions and the formula for the calculation of a_n and b_n.

RESULT 4.

(a) As $n \to \infty, nS_n \to_d$ exponential(1)

(b) For fixed *n*, the cdf of S_n is $F(s) = P(S_n \leq s) = 1 - \frac{1}{(1+s)^{n-1}}$

(c) The quantiles of S_n are: quantile(p) = $(1-p)^{\frac{-1}{n-1}} - 1$.

(d) The confidence statement in equation (8) is correct, with

$$a_n = (1-0.025)^{\frac{-1}{n-1}} - 1 \ , \ b_n = (1-0.975)^{\frac{-1}{n-1}} - 1 \ .$$

Table 1 gives values of na_n and nb_n for selected values of *n* between 2 and 50,000. Very large sample sizes like *n* = 50,000 are not empirically relevant but are included so that we can clearly see the asymptotic behavior of the entries. The first two columns of critical values give the results from a simulation with 1,000,000 replications, while the last two columns are calculated using Result 4. The simulation results and the analytical results are quite close, which is a good check on both. The convergence to the asymptotic values of na_n and nb_n, which are 0.0253 and 3.689, is reasonably fast, but in this case there is no reason to use the asymptotic values when the exact (finite *n*) values are available.

For quantiles corresponding to probabilities other than 0.025 and 0.975, we can use the formula in Result 4(c). For example, we might want to calculate the 0.05 and 0.95 quantiles to construct a 90% confidence interval in place of a 95% interval. Or, another motivation might be that we want to construct a one-sided confidence interval (upper bound only), which because of the skewness of the distribution of *u* will be shorter than a two-sided interval of equal confidence level. For example, with *n* = 25, a two-sided 95% confidence interval will have width $\frac{4.154-.0264}{25}\hat{\mu}$ = $0.1651\hat{\mu}$, whereas a one-sided (upper) 95% confidence interval will have width $0.1329\hat{\mu}$. (Here the quantile 0.1329 is calculated from part (c) of Result 4, with *p* = 0.95.)

Table 2 gives the results of a simulation to check the coverage rates of the confidence intervals using the values of a_n and b_n from Table 1 and the formula in equation (8). For purposes of comparison we also present the coverage rates from the intervals using the formula in equation (6), based on the true value of μ. These intervals would be infeasible with real data but can be constructed in a simulation setting. We also present the coverage rates from the intervals using the formula in equation (8) but the asymptotic values of na_n and nb_n (.0253 and 3.689). There are no surprises here. The infeasible intervals cover the correct fraction (95%) of the time, and so do the exact intervals using the tabulated finite sample values of na_n and nb_n. This should be true exactly, apart from the randomness of the simulation and rounding error in the calculations. The intervals that use the asymptotic values are fairly accurate in small samples, especially when $n \geq 25$. Still, the ability to construct exact finite sample confidence intervals is of theoretical interest and some practical importance.

3. OTHER DISTRIBUTIONS

In this section, we still consider the case of an intercept only (no regressors), but we allow for distributions of *u* other than exponential. That is, the model (1) still holds, and the distribution of the error *u* is known apart from unknown parameters. We will discuss in most detail the case that the errors are half-normal, that is, *u* is distributed as $N(0, \sigma^2)^+$, the non-negative truncation of a normal random variable.

The key to our treatment is that we will not calculate the MLE based on the actual distribution of *u*. Rather, we will continue to calculate exactly the same quantities as we did in the previous section. So, specifically, $\hat{\alpha}$ and $\hat{\mu}$ are still as defined in Result 1 (however, there is no

longer any claim that they are the MLE's), $\hat{u}_i = \hat{\alpha} - y_i$ as in the previous section, and S_n is still defined as in Result 3.

We will assume the following.

ASSUMPTION 1.

(a) $\mu \equiv E(u)$ exists.

(b) The density of *u* is strictly positive in an open neighborhood (0,*c*) for some $c > 0$.

(c) (scalability) *u* is distributed as a scalar θ times a random variable *u**, where the distribution of *u** is known. In particular, there are no unknown parameters in the distribution of *u**.

Assumption 1(a) ensures that a law of large numbers applies to *u*, and therefore to *y*, so that $\bar{y} \rightarrow_p (\alpha - \mu)$. Assumption 1(b) ensures that $u_{min} \rightarrow_p 0$ so that $\hat{\alpha}$ is a consistent estimate of α. Note that Assumption 1(b) does not rule out the possibility that the density of *u* equals zero at the point *u* = 0. It just ensures that $P(u < k)$ is strictly positive no matter how small *k* is. Therefore Assumptions 1(a) and 1(b) ensure that $\hat{\mu}$ is a consistent estimate of μ. Since our interest is not primarily in asymptotics, these assumptions could have been omitted. Assumption 1(c) is fundamental to the results of this section, however. It will hold for many commonly-assumed one parameter distributions for *u*. For example, it holds in the exponential case, since if *u* is distributed as exponential with mean μ, it can be written as μ times *u** where *u** is exponential with mean one. Similarly, this assumption holds in the half-normal case, since if *u* is distributed as $N(0, \sigma^2)^+$, then it can be expressed as σ times *u** where *u** is distributed as $N(0,1)^+$. It will not hold for all

one-parameter distributions, however; for example, it does not hold if u is distributed as $N(\mu, 1)^+$. And it will generally not hold for two-parameter distributions.

The point of Assumption 1(c) is that it ensures that the unknown scale factor θ cancels out of the distribution of S_n. We state this result formally.

RESULT 5. Suppose that Assumption 1(c) holds. Then S_n is pivotal; its distribution does not depend on α or θ.

Because S_n is pivotal, we can calculate its quantiles by simulation, for an arbitrary value of α (e.g. $\alpha = 0$) and an arbitrary value of the scaling parameter. For example, in the half-normal case, we can generate data $y_i = -u_i$ (i = 1,…,n) where the u_i are iid draws from $N(0,1)^+$. This data implies a value of S_n, and repeating this process many times reveals the quantiles of S_n .

Once we have the critical values a_n and b_n, we can construct confidence intervals as before, according to equation (8). The only difference is that the values of a_n and b_n are different for different distributions of u.

Table 3 gives the values of na_n and nb_n for the half-normal case, for selected values of n between 2 and 50,000, based on a simulation with 1,000,000 replications. This tabulation is similar to the simulation-based tabulation in Table 1 for the exponential case.

The last two columns of Table 3 give the coverage rates of the confidence intervals that use the asymptotic values of na_n and nb_n (.0398 and 5.7861) and the actual tabulated values, respectively. The coverage rates for the intervals that use the actual (finite sample) tabulated values are very close to the correct probability of 0.95, as they should be. The intervals that use the asymptotic values are reasonably accurate except when the sample size is very small.

Table 4 gives *n* times the quantiles of S_n for a variety of other probability levels that might be commonly used in practice. The values of na_n and nb_n that appear in Table 3 reappear in Table 4 in the columns corresponding to p = 0.025 and 0.975.

4. MODELS WITH REGRESSORS

We now consider the empirically relevant case that we have regressors (inputs) in our production function. That is, the model becomes

(9) $$y_i = \alpha + x_i'\beta - u_i \quad , \quad u_i \geq 0 \quad , \quad i = 1, 2, \ldots, n \quad ,$$

where x_i is a vector of inputs or functions (e.g. logarithms) of inputs. For the moment, we will assume that the u_i are exponential with mean μ. This assumption will be relaxed shortly.

Schmidt (1976) showed that, if the u_i are exponential, the MLE is the Aigner and Chu (1968) linear programming estimator. That is, the MLE's of α and β solve the problem:

(10) $$min_{\alpha,\beta} \sum_i |\alpha + x_i'\beta - y_i| \text{ subject to: } y_i \leq \alpha + x_i'\beta \text{ for all } i = 1,\ldots,n$$

Unfortunately, the statistical properties of this estimator are unknown. Thus we will consider instead the corrected ordinary least squares (COLS) estimator proposed by Winsten (1957) and Greene (1980). As intuitive motivation, we write the model as

(11) $$y_i = (\alpha - \mu) + x_i'\beta - (u_i - \mu) = \alpha^* + x_i'\beta - u_i^*$$

which now has an error with expectation zero and can be reasonably estimated by OLS (whether the errors are exponential or not). Let the estimates from OLS of y_i on intercept and x_i (i = 1,…,n) be denoted as $\hat{\alpha}^*$, $\hat{\beta}$, and let the OLS residuals be $e_i = y_i - \hat{\alpha}^* - x_i'\hat{\beta}$. Now define $\hat{\mu} = e_{max}$, the largest of the OLS residuals, and define $\hat{\alpha} = \hat{\alpha}^* + \hat{\mu}$. That is, we have shifted the OLS regression line upward (leaving its slope unchanged) by the amount of the largest residual, so that the line now goes through one point and lies above the other points. Finally, then, we define the

inefficiency estimates: $\hat{u}_i = \hat{\alpha} + x_i'\hat{\beta} - y_i$. So one of the $\hat{u}_i$ will equal zero, and the others will be positive.

The following result shows the sense in which this is a legitimate generalization of our approach of Section 2.

RESULT 6. In the intercept-only case, the exponential MLE and the COLS estimator are the same.

With regressors, the exponential MLE and the COLS estimator are not the same. We choose the COLS estimator, because we can construct exact confidence intervals based on these estimates, as we now proceed to show.

We now drop the assumption of exponential errors, and make the weaker assumption that the distribution of the errors satisfies Assumption 1, and therefore satisfies the scalability condition 1(c). As before we denote the scaling factor by θ. We then have the following result:

RESULT 7. Define $S_{n,i} = \frac{u_i - \hat{u}_i}{\hat{\mu}}$. Then $S_{n,i}$ is pivotal; its distribution does not depend on α, β or θ.

It is easy to see that

(12) $$\hat{u}_i - u_i = (\hat{\alpha} - \alpha) + x_i'(\hat{\beta} - \beta) .$$

Unlike in the case of intercept-only, this difference is not the same for every *i*, and its distribution depends on the value of the regressor matrix *X* and also on the specific value of x_i . As before, because $S_{n,i}$ is pivotal, we can calculate the quantiles of $S_{n,i}$ by simulation, generating the data for

the simulation using arbitrary values of α, β and θ. These quantiles will vary over i. Other than that, the procedure is essentially the same as before. We find (by simulation) the critical values $a_{n,i}$ and $b_{n,i}$ such that

(13) $$P\left(a_{n,i} \leq S_{n,i} \leq b_{n,i}\right) = 0.95$$

and this leads to the confidence interval:

(14) $$P\left(\hat{u}_i + a_{n,i}\hat{\mu} \leq u_i \leq \hat{u}_i + b_{n,i}\hat{\mu}\right) = 0.95 \ .$$

It is interesting that, unlike in the intercept-only case, the values of $S_{n,i}$ can now be of either sign. It will be the case that $S_{n,i} < 0$ when $u_i < \hat{u}_i$, that is, when, at a given value of x_i, the fitted line lies above the true line. However, because the fitted line and the true line can cross, this can be true for some (but not all) values of x_i. Therefore it is possible that $a_{n,i}$ can be negative for some values of i.

Table 5 presents the results of a simulation that verifies that the intervals (14) cover with the correct probability. The data are generated with an intercept and a dummy variable, and the errors are exponential. The dummy variable takes on the value of 1 for 60% of the observations and the value of 0 for 40% of the observations. (These are convenient fractions because all of our sample sizes are multiples of five.) Obviously inputs in a production function are not generally dummy variables. The choice of a dummy variable for the non-constant regressor is a matter of convenience. Although there are n observations on x_i, they only take on two different values, and so we have to calculate only two values of $a_{n,i}$ and $b_{n,i}$ (one set for the observations for which x_i = 0 and another set for the observations for which x_i = 1). Table 5 presents these two sets of $na_{n,i}$ and $nb_{n,i}$. We also can note that for observations with the same value of x_i , the estimation error $\hat{u}_i - u_i$ is the same, and so the intervals either cover or do not cover u_i for all values of i with that value of x_i. As a result we need to report only two (instead of n) coverage rates, one for the

observations with $x_i = 0$ and another for the observations with $x_i = 1$. These two coverage rates are given in the last two columns of Table 5. They are equal to 0.95 apart from the randomness of the simulation, as they should be.

5. UNKNOWN ERROR DISTRIBUTION

In this section we consider the case that the form of the distribution of *u* is unknown. In this case valid finite sample inference is no longer possible, but under certain assumptions asymptotically valid inference is possible using resampling methods. Interestingly, now the intercept-only case is fundamentally different from the case with regressors, because the rate of convergence of $\hat{u}_i - u_i$ differs in the two cases.

5A. The Intercept-Only Case

We consider first the intercept-only model of equation (2). As noted above, this is not just a matter of convenience of exposition; this case is different from the case with regressors. As before we are interested in a confidence interval for u_i , or equivalently a confidence interval for α, and we consider the estimate $\hat{\alpha} = y_{max}$.

In this case we will use subsampling (or the *m* out of *n* bootstrap) because the usual full-sample bootstrap is not valid. To see why the bootstrap fails in this case, let *B* be the number of bootstrap samples used and let *b* (= 1,…,*B*) be a single bootstrap sample of *n* observations chosen with replacement from $\{y_1, \ldots, y_n\}$. Let $\hat{\alpha}(b)$ be the estimate of α based on bootstrap sample *b*. The usual justification for the bootstrap is that asymptotically the distribution of $(\hat{\alpha}(b) - \hat{\alpha})$ is the same as that of $(\hat{\alpha} - \alpha)$. However, this result does not hold in the present case, or in general in many problems involving boundaries. The specific problem is that, while $\hat{\alpha}$ is strictly less than α with probability one, it will often be the case that $\hat{\alpha}(b) = \hat{\alpha}$. We will have $\hat{\alpha}(b) = \hat{\alpha}$

whenever the bootstrap sample contains y_{max}, and this will occur with probability $1-(1-\frac{1}{n})^n$. For large *n* this probability converges to 1 – 1/*e*, or approximately 0.63. So the bootstrap estimate will exactly equal the original estimate about 63% of the time, and this event has no counterpart in the relationship of the original estimate to the true parameter.

Politis and Romano (1994) and Politis, Romano and Wolf (1999) solved this problem by using subsamples (again denoted $b = 1,\ldots,B$) of size $m < n$. Formally we pick *m* such that, as $n \to \infty$, $m \to \infty$ but $m/n \to 0$. This is called "subsampling" if the subsamples are chosen without replacement, and it is called the "*m* out of *n* bootstrap" if they are chosen with replacement. (The usual bootstrap corresponds to $m = n$ and is ruled out by the assumption that $m/n \to 0$.) In the case of subsampling, the *B* subsamples can either be all possible subsamples of size *m* or they can be a random selection from the set of all subsamples of size *m*. If *m*/*n* is small, it will not matter much whether sampling is with or without replacement, so subsampling and the *m* out of *n* bootstrap will be similar. The point of making *m*/*n* small is that it reduces the probability that $\hat{\alpha}(b) = \hat{\alpha}$. For example, for the *m* out of *n* bootstrap with $m/n = 0.1$, for large *n* the probability that $\hat{\alpha}(b) = \hat{\alpha}$ equals $1 - e^{-0.1} = 0.095$, so the bootstrap estimate will equal the original estimate about 9.5% of the time, instead of 63% of the time. Conversely, the point of having *m* grow with *n* is so that asymptotics apply to the estimates from the subsamples in the same way as to the original estimate.

To make proper use of the subsampled estimates, we need to know the rate of convergence of the original estimate. Suppose that $\tau_n(\hat{\alpha} - \alpha)$ has a non-degenerate asymptotic distribution, where τ_n is a known function of *n*. (For many estimation problems, $\tau_n = \sqrt{n}$, and for the problem just discussed, generally $\tau_n = n$.) Then the basic result is that the asymptotic (large *n*) distribution of $\tau_m(\hat{\alpha}(b) - \hat{\alpha})$ is the same as the asymptotic distribution of $\tau_n(\hat{\alpha} - \alpha)$. See, for

example, Politis, Romano and Wolf (1999), Theorem 2.2.1, p. 43. Thus we approximate the distribution of $(\hat{\alpha} - \alpha)$ by the distribution of $(\frac{\tau_m}{\tau_n})(\hat{\alpha}(b) - \hat{\alpha})$. This leads us to an asymptotically valid confidence interval of the form

(15) $$P\big(\hat{\alpha} + a_{m,n} \le \alpha \le \hat{\alpha} + b_{m,n}\big) = 0.95 ,$$

where $a_{m,n}$ and $b_{m,n}$ are the 0.025 and 0.975 quantiles of the *B* values of $(\frac{\tau_m}{\tau_n})(\hat{\alpha} - \hat{\alpha}(b))$.

Finally, since $\hat{u}_i - u_i = \hat{\alpha} - \alpha$, we can use these same values of $a_{m,n}$ and $b_{m,n}$ for the confidence intervals

(16) $$P(\hat{u}_i + a_{m,n} \le u_i \le \hat{u}_i + b_{m,n}) = 0.95 .$$

In the intercept-only model of equation (2), the rate of convergence of $\hat{\alpha}$ depends on the behavior of the density of *u* near and at zero. We will make the following assumption.

ASSUMPTION 2.

(a) $\mu \equiv E(u)$ exists.

(b) The density of *u* is strictly positive in an open neighborhood (0,*c*) for some $c > 0$.

(c) $0 < f(0) < \infty$

Comparing this assumption to Assumption 1, we no longer assume the scalability property. However, we now assume that $0 < f(0) < \infty$. In this case, the following result implies the relevant asymptotic distribution for $\hat{\alpha}$.

RESULT 8. Let $u_1, \dots, u_n$ be a random sample from the distribution of a non-negative random variable with density $f(u)$. Suppose that Assumption 2 holds, so that $0 < f(0) < \infty$. Then $n \cdot u_{min} \rightarrow_d$ exponential(μ) with $\mu = 1/f(0)$.

Since $\hat{\alpha} - \alpha = -u_{min}$, Result 8 implies that $n(\hat{\alpha} - \alpha) \rightarrow_d$ – exponential(μ) with $\mu = 1/f(0)$. So to apply subsampling techniques, we choose $\tau_n = n$ in this case.

We note that, as an alternative to subsampling, we could base confidence intervals on the asymptotic distribution (exponential). However, this would require an estimate of $f(0)$, and we prefer to avoid the problems inherent in density estimation.

In the case that $f(0)$ equals zero [or infinity], the asymptotic distribution of u_{min} and the rate of convergence to this distribution depend on the rate of convergence of $f(0)$ to zero [or infinity]. Such cases could be empirically relevant but they will be very difficult to treat, and we therefore will not consider them here.

Table 6 gives the results of simulations to determine the coverage rates of confidence intervals constructed by subsampling and the *m* out of *n* bootstrap. We considered two sample sizes, *n* = 100 and 500. We use three different distributions: exponential, half-normal and uniform. The number of iterations was 10,000, and the number of subsamples (or bootstrap subsamples) used in each iteration was 1000. (In the case of subsampling, that is, sampling without replacement, this was a random sample of 1000 from the set of all possible subsamples.)

The exponential case is unusual because the finite sample distribution of the estimate is the same as the asymptotic distribution (i.e., exponential) no matter what the value of *m* is. So the requirement that *m* be large enough for asymptotics to be relevant is unnecessary. And, indeed, the confidence intervals are quite accurate (coverage is close to 0.95) so long as *m* is not too large. As expected, so long as *m* is small relative to *n*, subsampling and the *m* out of *n* bootstrap are quite similar.

For the other two distributions, the value of *m* is more important. If *m* is too small the coverage rates are too small. Presumably this occurs because the asymptotic distribution is a poor approximation to the finite sample distribution of the subsample estimates when the subsample size is very small. If *m* equals 10 or 20 when *n* = 100, or if *m* = 10, 20, 50 or 100 when *n* = 500, the coverage rates of the confidence intervals are reasonably close to 0.95. These values of *m* would be quite standard in the subsampling literature. For these values of *m*, once again there is little difference between subsampling and the *m* out of *n* bootstrap.

The general conclusion from these simulations is that subsampling (or the *m* out of *n* bootstrap) works quite well in the intercept-only case.

5B. The Model with Regressors

Now we turn to the empirically relevant case in which we have regressors (inputs) in our production function. The model is as given in equation (9) of section 4. However, now we do not assume that the distribution of the error is known. We simply assume that the distribution of *u* satisfies Assumption 2.

Interestingly, the rate of convergence of $\hat{u}_i - u_i$ is different in the case with regressors than in the intercept-only case. To see why this is true, we begin with the following result.

RESULT 9. $\hat{u}_i - u_i = x_i'(\hat{\beta} - \beta) - min_{j=1,\ldots,n}(u_j + x_j'(\hat{\beta} - \beta))$

Since $\sqrt{n}(\hat{\beta} - \beta)$ has a (normal) asymptotic distribution, the same is true of $\sqrt{n}(\hat{u}_i - u_i)$. That is, it is the slower-converging portion of the right hand side of the expression in Result 9 that determines the rate of convergence. Therefore we take $\tau_n = \sqrt{n}$ for subsampling.

It may be worthwhile to be explicit about the nature of our subsampling and the construction of the resulting confidence intervals. We sample (*y*,*x*) pairs, so that subsample *b* is a set of *m* observations randomly drawn from $\{(y_1, x_1), \ldots, (y_n, x_n)\}$, either without replacement ("subsampling") or with replacement ("*m* out of *n* bootstrap"). Let $\hat{\alpha}(b)$ and $\hat{\beta}(b)$ be the estimates of α and β in subsample *b*. These imply estimates $\hat{u}_i(b) = \hat{\alpha}(b) + x_i'\hat{\beta}(b) - y_i$ for all *i*. That is, we obtain $\hat{u}_i(b)$ whether or not observation *i* was one of the *m* observations that went into subsample *b*. Now $a_{m,n,i}$ and $b_{m,n,i}$ are defined as the 0.025 and 0.975 quantiles of the *B* values of $(\hat{u}_i - \hat{u}_i(b))$, and they yield the confidence intervals:

(17) $$P(\hat{u}_i + a_{m,n,i} \leq u_i \leq \hat{u}_i + b_{m,n,i}) = 0.95 .$$

Tables 7, 8 and 9 give results of simulations to determine the coverage rates of confidence intervals constructed by subsampling and the *m* out of *n* bootstrap. As in the intercept-only case, we considered two sample sizes, *n* = 100 and 500, and we used three different distributions: exponential, half-normal and uniform. The number of iterations was 10,000, and the number of subsamples (or bootstrap subsamples) used in each iteration was 1000. Our model contains intercept and a single regressor (*x*). For *n* = 100 we used 100 different uniformly spaced values of *x*: *x* = 0, 0.1, 0.2,0.3,…, 9.9. For *n* = 500 we used these same 100 values, repeated five times each.

Table 7 gives the results for *x* = 0, the smallest value of *x*. (The results for the largest value, *x* = 9.9, are very similar.) These results are quite good. Consider subsampling first. For *n* = 100 and *m* = 10 or 20, we have coverage rates in the range of 90-95% and the results are only a little worse than in the intercept-only case. For *n* = 500 and *m* = 10, 20, 50 or 100, the results are better, and in fact are better than in the intercept-only case. For the *m* out of *n* bootstrap, and with *m* in the same range as just discussed, the results are quite similar to those for subsampling.

A striking difference between these results and the results for the intercept-only case is that the full-sample bootstrap ($m = n$) now yields approximately correct coverage rates. We will comment more on this below.

The results for $x = 2.4$, the 25th percentile value of x, are given in Table 8, while Table 9 gives the results for $x = 4.9$, the median value of x. We will focus on the results in Table 9 because the results in Table 8 are intermediate between those in Tables 7 and 9. For subsampling and for the m out of n bootstrap with relatively small values of m, the results are clearly worse than for the case of $x = 0$ (Table 7). Coverage rates are now in the range of 80-90% as opposed to 90-95%. This result was unexpected. It is true that the effect of estimation error in β is smallest at the sample mean, and that this is the component of estimation error for u_i that converges at rate $\sqrt{n}$. However, the term $(\hat{\beta} - \beta)$ does not cancel from Result 9 even when $x_i = \bar{x}$ and so the convergence rate remains $\sqrt{n}$ even for $x_i = \bar{x}$. The finite sample relevance of this convergence rate may simply be weaker at or near the sample mean.

The most interesting aspect of the results in Tables 7, 8 and 9 is that the ordinary (full sample, $m = n$) bootstrap works quite well regardless of the value of x. We are not aware of any proof of the asymptotic validity of the bootstrap that would apply to this case, but our simulations certainly support the notion that it is valid. We intend to pursue this question in a separate paper. In any case, as a practical matter, based on our simulations the bootstrap would be the preferred method.

6. CONCLUDING REMARKS

The main purpose of this paper was to resolve an intellectual puzzle. For the linear deterministic frontier model with known error distribution, valid methods of inference on the

inefficiencies were not available, even though asymptotically valid methods of inference were available for more complicated models. The linear deterministic frontier model with known error distribution makes very strong assumptions and so we ought to be able to do at least as well as in models that make weaker assumptions. We show that this is true. Exact (finite sample) inference is possible if the error distribution is known up to a scaling parameter.

A secondary purpose of the paper was to investigate the case that the error distribution is not known. Now asymptotically valid inference should be possible using resampling (subsampling or bootstrapping) methods. Interestingly, the intercept-only case differs from the case with regressors (which is the empirically relevant case), because the rate of convergence of the inefficiency estimates is different with regressors than without. The bootstrap is known not to be valid in the intercept-only case, whereas it performs well in simulations in the case with regressors. The theoretical basis for this good performance remains as a topic for further research.

APPENDIX

Proof of Result 1. This result follows from Lawless (1982), unnumbered equations on p. 127. However, for the convenience of the reader we give a brief proof. The density of *y* is $f(y) = \left(\frac{1}{\mu}\right) \exp\left(-\frac{\alpha - y}{\mu}\right)$ for $\alpha \geq y$ and so the log-likelihood is $lnL = n\, ln\, \mu - \frac{1}{\mu}\sum_i(\alpha - y_i)$ $= n\, ln\, \mu - \frac{n}{\mu}(\alpha - \bar{y})$ for $\alpha \geq y_i$ for all $i = 1,\ldots,n$. Maximizing partially with respect to μ yields $\mu = \alpha - \bar{y}$, and substituting this into *lnL* gives the concentrated likelihood:

(A1) $$lnL^* = -n \ln(\alpha - \bar{y}) - n \text{ for } \alpha \geq y_i \text{ for all } i = 1,\ldots,n.$$

Thus we minimize $\ln(\alpha - \bar{y})$, or equivalently minimize α, subject to the constraint. This yields $\hat{\alpha} = y_{max}$ and then $\hat{\mu} = \hat{\alpha} - \bar{y}$.

Proof of Result 2. We use the results that $E(\bar{u}) = \mu$ and $E(u_{min}) = \mu/n$. Therefore $E(\bar{y}) = \alpha - \mu$ and $E(y_{max}) = \alpha - \mu/n$. Then $E(\tilde{\alpha}) = \frac{n}{n-1}\left(\alpha - \frac{\mu}{n}\right) - \frac{1}{n-1}(\alpha - \mu) = \alpha$ and similarly for $\tilde{\mu}$. In terms of the algebraic relationship of the unbiased and biased estimates, the relationship between $\tilde{\mu}$ and $\hat{\mu}$ is obvious, and for the relationship between $\tilde{\alpha}$ and $\hat{\alpha}$ we have

(A2) $$\tilde{\alpha} = \frac{n}{n-1} y_{max} - \frac{1}{n-1}\bar{y} = y_{max} + \frac{1}{n-1}(y_{max} - \bar{y}) = \hat{\alpha} + \frac{1}{n-1}\hat{\mu}$$

Proof of Result 3. This result is pointed out by Lawless (1982), p. 128, but not explicitly proved. We can write

(A3) $$S_n = u_{min} / (\bar{u} - u_{min})$$

which does not involve α. Its distribution also does not depend on μ because μ is a scale factor in both numerator and denominator and cancels from the ratio.

Proof of Result 4.

(a) Similarly to (A3), write $nS_n = nu_{min} / (\bar{u} - u_{min})$. Then nu_{min} is distributed as exponential(μ), while $\bar{u} \rightarrow_p \mu$ and $u_{min} \rightarrow_p 0$, and so $nS_n \rightarrow_d$ exponential(μ) / μ = exponential(1).

(b) Invert the quantile function in part (c) to get the distribution function, that is, solve $s = (1-p)^{\frac{-1}{n-1}} - 1$ for p in terms of s.

(c) It follows from Lawless (1982), equation (3.5.3) that $(n-1)S_n$ is distributed as $F_{2,2(n-1)}$. The quantile is then given in his equation (3.5.4).

(d) Follows immediately from (c).

Proof of Result 5. The proof is essentially the same as the proof of Result 3. Equation (A3) still holds, and it does not involve α, and the effect of the scale factor θ cancels from the ratio in (A3).

Proof of Result 6. Consider the equivalent of equation (11) in the case of no regressors: $y_i = \alpha^* + u_i^*$ where $\alpha^* = \alpha - \mu$. This is a model with only an intercept, and so the OLS estimator is the sample mean, that is, $\hat{\alpha}^* = \bar{y}$. Then the residuals are $e_i = y_i - \bar{y}$, and $\hat{\mu} = e_{max} = y_{max} - \bar{y}$. Finally, the COLS estimate is $\hat{\alpha} = \hat{\alpha}^* + \hat{\mu} = \bar{y} + (y_{max} - \bar{y}) = y_{max}$, which is the same as the exponential MLE.

Proof of Result 7. The proof follows the same lines as the proofs of Results 3 and 5. We start with the numerator, $\hat{u}_i - u_i = (\hat{\alpha} - \alpha) + x_i'(\hat{\beta} - \beta)$. The distributions of $\hat{\alpha} - \alpha$ and $\hat{\beta} - \beta$ do not depend on the true values of α and β, and the whole expression is linear in the errors (u's) and therefore has the same scale as θ. Similar statements apply to the denominator. Each of the individual residuals has a distribution that does not depend on the true values of α and β, and has the same scale as θ. The same is then true of $\hat{\mu} = e_{max}$. Finally, the scale factor θ cancels from the numerator and denominator.

Proof of Result 8. This result can be derived as a consequence of Corollary 1.3.2 of Galambos (1978), p. 11. However, a direct proof is simpler. Let $F(u)$ be the cdf corresponding to $f(u)$. Then note that, for fixed $c > 0$, as $n \to \infty$, since $F(0) = 0$,

(A4) $$nF\left(\frac{c}{n}\right) = \frac{F(\frac{c}{n})}{c/n} \cdot c \to f(0)c$$

Now we evaluate

$$P(n \cdot u_{min} > c) = P\left(u_{min} > \frac{c}{n}\right) = [1 - F\left(\frac{c}{n}\right)]^n,$$

so that

$$lnP(n \cdot u_{min} > c) = n \cdot \ln\,[1 - F\left(\frac{c}{n}\right)].$$

But $\ln(1 + x) = x + o(x)$, so

$$n \cdot \ln\,[1 - F\left(\frac{c}{n}\right)] = -n \cdot F\left(\frac{c}{n}\right) + o\left[n \cdot F\left(\frac{c}{n}\right)\right] \to -f(0)c\ ,$$

using (A4). So

(A5) $$P(n \cdot umin > c) = \exp\,[ln\,P(n \cdot umin > c)] \rightarrow e^{-f(0)c}.$$

This is the survivor function for the exponential distribution with mean $\mu = 1/f(0)$.

Proof of Result 9. We begin with the expression given in equation (12), $\hat{u}_i - u_i = (\hat{\alpha} - \alpha) + x_i'(\hat{\beta} - \beta)$. Now we seek to substitute out $(\hat{\alpha} - \alpha)$. We have $\hat{\alpha} = \hat{\alpha}^* + \hat{\mu}$ where $\hat{\alpha}^* = \bar{y} - \bar{x}'\hat{\beta}$ and $\hat{\mu} = max_j e_j$. But $e_j = \left(y_j - \bar{y}\right) - (x_j - \bar{x})'\hat{\beta}$. Therefore $\hat{\mu} = max_j(y_j - x_j'\hat{\beta}) - (\bar{y} - \bar{x}'\hat{\beta})$ and $\hat{\alpha} = max_j(y_j - x_j'\hat{\beta})$. Now $y_j - x_j'\hat{\beta} = \alpha - u_j - x_j'(\hat{\beta} - \beta)$ and therefore $\hat{\alpha} - \alpha = max_j(-u_j - x_j'\left(\hat{\beta} - \beta\right)) = -min_j(u_j + x_j'\left(\hat{\beta} - \beta\right))$. With this substitution we obtain Result 9.

REFERENCES

Aigner, D.J. and S.F. Chu (1968), "On Estimating the Industry Production Function," *American Economic Review* 58, 826-839.

Aigner, D.J., C.A.K. Lovell and P. Schmidt (1977), "Formulation and Estimation of Stochastic Frontier Production Function Models," *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.

Galambos, J. (1978), *The Asymptotic Theory of Extreme Order Statistics*, New York: Wiley.

Greene, W.H. (1980), "Maximum Likelihood Estimation of Econometric Frontier Functions," *Journal of Econometrics* 13, 27-56.

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

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

Kneip, A., B.U. Park and L. Simar (1998), "A Note on the Convergence of Nonparametric DEA Efficiency Measures," *Econometric Theory* 14, 783-793.

Kneip, A., L. Simar and P.W. Wilson (2008), "Asymptotics and Consistent Bootstraps for DEA Estimators in Non-Parametric Frontier Models," *Econometric Theory* 24, 1663-1697.

Kumbhakar, S.C. and C.A.K. Lovell (2000), *Stochastic Frontier Analysis*, Cambridge: Cambridge University Press.

Lawless, J.F. (1982), *Statistical Models and Methods for Lifetime Data*, New York: Wiley.

Meeusen, W. and J. van den Broeck (1977), "Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error," *International Economic Review* 18, 435-444.

Park, B.U., L. Simar and C. Wiener (2000), "The FDH Estimator for Productivity Efficiency Scores: Asymptotic Properties," *Econometric Theory* 16, 855-877.

Park, B.U., S.O. Jeong and L. Simar (2010), "Asymptotic Distribution of Conical-Hull Estimators of Directional Edges," *Annals of Statistics* 38, 1320-1340.

Politis, D. and J.P. Romano (1994), "Large Sample Confidence Regions Based on Subsamples under Minimal Assumptions," *Annals of Statistics* 22, 2031-2050.

Politis, D., J.P. Romano and M. Wolf (1999), *Subsampling*, New York: Springer.

Richmond, J. (1974), "Estimating the Efficiency of Production," *International Economic Review* 15, 515-521.

Schmidt, P. (1976), "On the Statistical Estimation of Parametric Frontier Production Functions," *Review of Economics and Statistics* 58, 238-239.

Simar, L. and P.W. Wilson (1998), "Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Nonparametric Frontier Models," *Management Science* 44, 49-61.

Simar, L. and P.W. Wilson (2000A), "A General Methodology for Bootstrapping in Nonparametric Frontier Models." *Journal of Applied Statistics* 27, 779-802.

Simar, L. and P.W. Wilson (2000B), "Statistical Inference in Nonparametric Frontier Models: The State of the Art," *Journal of Productivity Analysis* 13, 49-78.

Simar, L. and P.W. Wilson (2009), "Inference by Subsampling in Nonparametric Frontier Models," unpublished manuscript.

Simar, L. and P.W. Wilson (2010), "Inferences from Cross-Sectional Stochastic Frontier Models," *Econometric Reviews* 29, 62-98.

Winsten, C.B. (1957), "Discussion on Mr. Farrell's Paper," *Journal of the Royal Statistical Society, Series A* 120, 282-284.

TABLE 1

Values of na_n and b_n , intercept only, exponential errors

	By simulation		Analytic formula	
n	na_n	nb_n	na_n	nb_n
2	0.0510	78.008	0.0513	78.000
3	0.0380	15.960	0.0382	15.974
5	0.0317	7.593	0.0317	7.574
10	0.0282	5.070	0.0282	5.066
25	0.0263	4.143	0.0264	4.154
50	0.0256	3.898	0.0258	3.909
100	0.0258	3.783	0.0256	3.796
200	0.0254	3.738	0.0254	3.742
500	0.0255	3.706	0.0254	3.710
1000	0.0252	3.689	0.0253	3.699
2000	0.0253	3.684	0.0253	3.694
5000	0.0256	3.692	0.0253	3.691
10,000	0.0253	3.688	0.0253	3.690
25,000	0.0251	3.700	0.0253	3.689
50,000	0.0253	3.695	0.0253	3.689

TABLE 2

Coverage rates of 95% confidence intervals

Intercept only, exponential case

n	Infeasible using true μ (equation (6))	Using asymptotic values of na_n and nb_n	Using exact values of na_n and nb_n
2	0.9498	0.6360	0.9501
3	0.9504	0.7822	0.9502
5	0.9500	0.8700	0.9500
10	0.9501	0.9182	0.9504
25	0.9503	0.9398	0.9503
50	0.9496	0.9445	0.9496
100	0.9498	0.9474	0.9493
200	0.9498	0.9485	0.9497
500	0.9503	0.9497	0.9497
1000	0.9503	0.9500	0.9500
2000	0.9499	0.9498	0.9498
5000	0.9503	0.9502	0.9502
10,000	0.9501	0.9501	0.9501
25,000	0.9499	0.9498	0.9498
50,000	0.9499	0.9499	0.9499

TABLE 3

Values of na_n and nb_n and coverage rates of 95% confidence intervals

Intercept only, half-normal case

n	na_n	nb_n	Coverage, using asymptotic values of na_n and nb_n	Coverage, using exact values of na_n and nb_n
2	0.0794	100.46	0.6669	0.9502
3	0.0602	21.311	0.8049	0.9494
5	0.0498	10.685	0.8842	0.9502
10	0.0442	7.5334	0.9245	0.9501
25	0.0414	6.3987	0.9410	0.9499
50	0.0407	6.0709	0.9456	0.9497
100	0.0397	5.9223	0.9480	0.9502
200	0.0396	5.8682	0.9488	0.9503
500	0.0398	5.8301	0.9495	0.9502
1000	0.0401	5.8168	0.9498	0.9501
2000	0.0399	5.7968	0.9496	0.9497
5000	0.0396	5.7870	0.9498	0.9500
10,000	0.0398	5.8039	0.9497	0.9500
25,000	0.0403	5.7956	0.9501	0.9499
50,000	0.0398	5.7861	0.9497	0.9497

TABLE 4

Quantiles of nS_n

Intercept only, half-normal case

n	.01	.025	.05	.10	.90	.95	.975	.99
2	0.0315	0.0794	0.1646	0.3426	23.477	49.252	100.46	257.56
3	0.0239	0.0602	0.1228	0.2537	8.9244	14.084	21.311	35.575
5	0.0196	0.0498	0.1009	0.2085	5.6658	7.9700	10.685	14.917
10	0.0174	0.0442	0.0903	0.1855	4.4134	5.9111	7.5334	9.7980
25	0.0161	0.0414	0.0840	0.1726	3.8940	5.1259	6.3987	8.0599
50	0.0158	0.0407	0.0818	0.1688	3.7528	4.9035	6.0709	7.6149
100	0.0159	0.0397	0.0812	0.1671	3.6859	4.8052	5.9223	7.4262
200	0.0157	0.0396	0.0806	0.1663	3.6577	4.7700	5.8682	7.3539
500	0.0160	0.0398	0.0811	0.1662	3.6301	4.7293	5.8301	7.2832
1000	0.0159	0.0401	0.0807	0.1658	3.6245	4.7223	5.8168	7.2585
2000	0.0158	0.0399	0.0806	0.1660	3.6231	4.7208	5.7968	7.2423
5000	0.0156	0.0396	0.0805	0.1656	3.6174	4.7063	5.7870	7.2341
10,000	0.0159	0.0398	0.0809	0.1658	3.6173	4.7084	5.8039	7.2646
25,000	0.0160	0.0403	0.0814	0.1658	3.6251	4.7036	5.7956	7.2279
50,000	0.0159	0.0398	0.0805	0.1656	3.6144	4.7044	5.7861	7.2257

TABLE 5

Values of $na_{n,i}$ and $nb_{n,i}$ and coverage rates of 95% confidence intervals

Model with dummy variable regressor, exponential errors

n	$na_{n,i}$ $x = 0$	$nb_{n,i}$ $x = 0$	$na_{n,i}$ $x = 1$	$nb_{n,i}$ $x = 1$	Coverage $x = 0$	Coverage $x = 1$
5	-4.2177	20.864	-3.6212	14.444	0.9501	0.9501
10	-7.2147	10.448	-6.4286	7.3407	0.9499	0.9500
25	-13.686	7.9517	-12.567	5.6138	0.9500	0.9497
50	-21.394	7.4877	-20.035	5.2313	0.9502	0.9492
100	-32.697	7.3202	-31.071	5.0593	0.9502	0.9499
200	-48.733	7.2759	-47.052	4.9760	0.9501	0.9504
500	-81.479	7.3282	-79.323	4.9734	0.9496	0.9495
1000	-118.15	7.3296	-115.98	4.9623	0.9495	0.9508
2000	-170.57	7.3848	-167.65	4.9717	0.9502	0.9498
5000	-274.32	7.4184	-272.38	4.9631	0.9512	0.9497

TABLE 6

Coverage rates of 95% confidence intervals
Intercept only

$n = 100$

	Subsampling			*m* out of *n* bootstrap		
m	**Exponential**	**Half-normal**	**Uniform**	**Exponential**	**Half-normal**	**Uniform**
1	0.9438	0.8090	0.6001	0.9486	0.8095	0.5986
2	0.9523	0.8705	0.7888	0.9526	0.8714	0.7909
3	0.9632	0.9066	0.8677	0.9683	0.9052	0.8718
4	0.9670	0.9238	0.9066	0.9711	0.9298	0.9036
5	0.9690	0.9332	0.9181	0.9704	0.9383	0.9196
10	0.9622	0.9366	0.9357	0.9643	0.9491	0.9415
20	0.9442	0.9327	0.9274	0.9629	0.9538	0.9507
50	0.8604	0.8473	0.8528	0.9414	0.9336	0.9532
100				0.8778	0.8777	0.8746

$n = 500$

	Subsampling			*m* out of *n* bootstrap		
m	**Exponential**	**Half-normal**	**Uniform**	**Exponential**	**Half-normal**	**Uniform**
1	0.9509	0.8058	0.5992	0.9506	0.8113	0.6034
2	0.9486	0.8753	0.7871	0.9499	0.8705	0.7874
3	0.9530	0.8990	0.8612	0.9524	0.8987	0.8560
4	0.9508	0.9069	0.8817	0.9516	0.9114	0.8895
5	0.9554	0.9191	0.9018	0.9535	0.9200	0.9044
10	0.9558	0.9434	0.9329	0.9547	0.9367	0.9358
20	0.9698	0.9591	0.9584	0.9714	0.9606	0.9627
50	0.9662	0.9583	0.9619	0.9702	0.9640	0.9647
100	0.9469	0.9445	0.9459	0.9620	0.9641	0.9590
200	0.8956	0.8980	0.8952	0.9431	0.9473	0.9416
300	0.8058	0.7979	0.8054	0.9285	0.9253	0.9345
400	0.6870	0.6986	0.6828	0.9080	0.9083	0.9086
500				0.8808	0.8755	0.8800

TABLE 7

x = 0.0 (smallest x)

Coverage rates of 95% confidence intervals

n = 100

	Subsampling			m out of n bootstrap		
m	Exponential	Half-normal	Uniform	Exponential	Half-normal	Uniform
5	.9468	.9353	.9172	.9563	.9328	.9227
10	.9397	.9215	.9011	.9428	.9263	.9095
20	.9352	.9249	.9201	.9544	.9445	.9278
50	.8756	.8768	.8752	.9631	.9527	.9504
100				.9435	.9389	.9368

n = 500

	Subsampling			m out of n bootstrap		
m	Exponential	Half-normal	Uniform	Exponential	Half-normal	Uniform
5	.9394	.9150	.9041	.9386	.9187	.8996
10	.9346	.9117	.8819	.9438	.9154	.8894
20	.9484	.9267	.9102	.9477	.9291	.9141
50	.9548	.9455	.9290	.9598	.9446	.9303
100	.9493	.9484	.9372	.9663	.9605	.9544
200	.9117	.9129	.9151	.9652	.9601	.9604
300	.8142	.8246	.8304	.9570	.9572	.9543
400	.6498	.6588	,6678	.9479	.9426	.9477
500				.9322	.9366	.9400

TABLE 8

x = 2.4 (25th percentile x)

Coverage rates of 95% confidence intervals

n = 100

	Subsampling			*m* out of *n* bootstrap		
m	**Exponential**	**Half-normal**	**Uniform**	**Exponential**	**Half-normal**	**Uniform**
5	.9061	.8912	.8804	.9178	.8926	.8865
10	.9131	.8911	.8762	.9150	.8985	.8863
20	.9159	.9086	.9096	.9349	.9287	.9185
50	.8534	.8690	.8719	.9456	.9430	.9480
100				.9309	.9250	.9232

n = 500

	Subsampling			*m* out of *n* bootstrap		
m	**Exponential**	**Half-normal**	**Uniform**	**Exponential**	**Half-normal**	**Uniform**
5	.8930	.8577	.8473	.8917	.8669	.8369
10	.9084	.8760	.8407	.9151	.8812	.8461
20	.9333	.9050	.8826	.9321	.9053	.8878
50	.9438	.9310	.9125	.9503	.9332	.9144
100	.9314	.9351	.9242	.9556	.9518	.9441
200	.8954	.8994	.9004	.9509	.9506	.9515
300	.8131	.8375	.8285	.9420	.9457	.9480
400	.6713	.6777	.6880	.9422	.9352	.9396
500				.9279	.9263	.9272

TABLE 9

x = 4.9 (median x)

Coverage rates of 95% confidence intervals

n = 100

	Subsampling			m out of n bootstrap		
m	Exponential	Half-normal	Uniform	Exponential	Half-normal	Uniform
5	.7657	.7666	.7795	.7779	.7738	.7859
10	.8069	.7940	.8014	.8207	.8044	.8084
20	.8529	.8614	.8669	.8805	.8771	.8807
50	.8255	.8448	.8607	.9310	.9323	.9367
100				.9328	.9209	.9195

n = 500

	Subsampling			m out of n bootstrap		
m	Exponential	Half-normal	Uniform	Exponential	Half-normal	Uniform
5	.7425	.6846	.6629	.7322	.6906	.6545
10	.7849	.7293	.6826	.7939	.7418	.6886
20	.8512	.8059	.7751	.8515	.8061	.7760
50	.8884	.8650	.8471	.8992	.8753	.8507
100	.8882	.8881	.8829	.9210	.9109	.9055
200	.8724	.8734	.8815	.9352	.9358	.9390
300	.8022	.8147	.8158	.9352	.9345	.9424
400	.6587	.6828	.6918	.9410	.9348	.9407
500				.9346	.9301	.9315