Empirical Economics (2000) 25:189–208

EMPIRICAL
ECONOMICS

Is it efficient to analyse efficiency rankings?

Uwe Jensen*

Institut für Statistik und Ökonometrie, Christian-Albrechts-Universität, Olshausenstr. 40, D-24118 Kiel, Germany (e-mail: Jensen@stat-econ.uni-kiel.de)

First version received: May 1998/final version accepted: July 1999

Abstract. When production functions are estimated as frontier functions, the deviations from the frontier can be interpreted as individual inefficiency estimates. Unfortunately, it has recently been shown that efficiency differences across individuals are very often statistically insignificant. In this paper, we will analyse the consequences of the consideration of confidence statements for the reliability of efficiency rankings. The stochastic frontier and confidence intervals derived by Horrace and Schmidt are compared to the COLS approach and bootstrap confidence intervals. The membership function is proposed as a simple Monte-Carlo approximation for the probability for an individual to be the most efficient in the sample.

Key words: Bootstrap, confidence intervals, frontiers, inefficiency, membership function, ranking

JEL classification: C2, C4, C6, J3

1. Introduction

It is well-known since Farrell (1957) and Aigner and Chu (1968) that production (or cost) functions should be estimated as frontier functions because – consistent with microeconomic theory – maximum possible output for given inputs then can be estimated. In this approach, the deviations from the frontier can be interpreted as individual inefficiency estimates. Applications include

* The author would like to thank two referees for their very detailed comments on the first two versions of this paper which led to improvements in all aspects, readability, correctness and precision. Thanks for helpful comments address to the participants of the 'Whitsun meeting' of the German Statistical Association in München and of the ESEM in Berlin. All residual errors are my own.

the estimation of dairy farm efficiency by Battese and Coelli (1988), the estimation of electric-utility efficiency by Reifschneider and Stevenson (1991), the estimation of bank efficiency by Caudill et al. (1995) or the estimation of research efficiency by Backes-Gellner (1989). In all these (and very many other) studies, the respective vector of efficiency estimates is carefully analysed and explained – in most cases unfortunately without any confidence statements.

But, in recent years, the required results have been published. Taube (1988) and Horrace and Schmidt (1996) derive confidence intervals for individual inefficiency estimates in the stochastic frontier model. Simar (1992) introduces the bootstrap approach to frontier models in order to analyse the sensitivity of the efficiency measures. Hall et al. (1995) apply the bootstrap approach to panel data. Ferrier and Hirschberg (1997) and Simar and Wilson (1998) give bootstrap confidence intervals for individual efficiency scores in DEA models. And the extent of uncertainty about individual inefficiency has been emphasized as well. Simar, considering FDH models with the help of bootstrap confidence intervals, notes that '*the rankings … are certainly to be taken with care*' (1992, p. 191). Horrace and Schmidt find out that – in stochastic frontier models – efficiency differences across individuals are very often statistically insignificant.

Of course, these results can have a disastrous effect on the profitable applicability of frontier models. If – e.g. in stochastic frontier models – analysing individual efficiency vectors only means having trouble with indirectly constructed, inconsistent and imprecise estimates, there are few arguments for undertaking this effort. In this paper, we will analyse the consequences of the consideration of confidence intervals for the reliability of efficiency rankings in parametric frontier models. It is common practice – not only in the application of frontier functions – to compile rankings without guarding them by confidence statements. E.g. in Germany, many efficiency rankings of this kind have been prepared for the evaluation of the research and teaching quality of universities, a highly sensitive point of present-day controversy. What is the value of those rankings if they are based on nothing more than sampling error?

The second section summarizes the necessary theory on frontier functions, and it gives a critical review on the two parametric frontier models – stochastic frontier and COLS – analysed in the subsequent sections. Section 3 shortly reviews the illustrating empirical example. Section 4 presents the dramatic consequences of confidence statements on the explanatory power of efficiency rankings. But if one concentrates on e.g. calculating the probability for an individual to be the most efficient in the sample, uncertainty is reduced to some extent. The membership function, a simple Monte-Carlo approximation for this probability, is introduced. The subsequent section analyses efficiency rankings derived from a COLS approach (which does not allow for statistical noise). A procedure for the calculation of bootstrap confidence intervals for individual efficiency estimates is proposed, and the membership function turns out to be useful again. Conclusions are drawn in the last section.

2. Frontier functions

This section summarizes the theory on frontier functions that is needed in the following.

Aigner and Chu (1968) develop the first parametric frontier model for the estimation of a Cobb-Douglas production function

$$y_i = \alpha + \sum_{j=1}^{k} \beta_j x_{ij} - u_i, \quad u_i \geq 0, \quad i = 1, \ldots, n \tag{1}$$

where y_i is output in logs, x_{ij} are inputs in logs and $\hat{y}_i$ is estimated maximum possible output for given inputs (in line with microeconomic theory) of individual no. i. The technical inefficiency $TE_i = \exp(-u_i)$ of individual no. i can be estimated via $\hat{u}_i$.

Thus, the task is to estimate a function lying 'on top of the data cloud'. See the surveys in Førsund et al. (1980) or Greene (1993) for the various theoretical and practical problems arising in this field. We will only mention three of them:

- Of course, it is impossible to determine the theoretical frontier giving the 'true' maximum possible output. The estimation depends on the finite set of individuals in the sample giving a so-called 'best-practice frontier' which is biased downwards.
- It will be seen in this section that it is a problem to develop techniques for estimating frontiers with the ability to estimate individual inefficiency reliably.
- Schmidt (1976) was the first to note that estimating frontier models often means running into problems of irregularity because the range of y_i depends on the parameters to be estimated:

$$y_i \in \left(-\infty, \alpha + \sum_{j=1}^{k} \beta_j x_{ij}\right) \tag{2}$$

In the following, we will introduce the two most popular parametric frontier approaches (stochastic frontier and COLS).

2.1. *Corrected OLS (COLS) approach*

The 'corrected OLS (COLS) method' is based on a very simple idea. With $E(u_i) = \mu \geq 0$, we can transform (1) to

$$y_i = \tilde{\alpha} + \sum_{j=1}^{k} \beta_j x_{ij} - \tilde{u}_i, \quad i = 1, \ldots, n \tag{3}$$

with the centered constant and the centered residuals

$$\tilde{\alpha} = \alpha - \mu \quad \text{and} \quad \tilde{u}_i = u_i - \mu \tag{4}$$

where the new error term $\tilde{u}$ has zero mean. With the assumptions of the standard regression model (except normality), OLS now provides consistent and BLU estimates for $(\tilde{\alpha}, \beta_1, \ldots, \beta_k)$. In the second (correction) step of the COLS method, the OLS function is simply shifted up to

$$\hat{\alpha} = \hat{\hat{\alpha}} - \min_{i=1,\ldots,n} \{\hat{\hat{u}}_i\} \tag{5}$$

(see Winsten (1957)) implying

$$\hat{u}_i = \hat{\hat{u}}_i - \min_{i=1,\ldots,n} \{\hat{\hat{u}}_i\} \tag{6}$$

(see Lovell (1993) for the 'modified OLS' or MOLS). The β_j retain their optimal OLS properties, and Greene (1980) shows the consistency of $\hat{\alpha}$. Bootstrap confidence intervals for the COLS constant can be constructed applying the procedure by Hall et al. (1995) developed for stochastic panel frontiers.

If the untransformed output $Y_i = \exp(y_i)$ is modeled instead of y_i, equation (1) becomes

$$Y_i = \exp(\alpha) \cdot \prod_{j=1}^{k} \exp(\beta_j x_{ij}) \cdot \exp(-u_i), \quad u_i \geq 0, \quad i = 1, \ldots, n \tag{7}$$

Then, point estimates for the untransformed individual technical efficiency $TE_i = \exp(-u_i)$ are

$$\widehat{TE}_i = \exp(-\hat{u}_i) = \exp\left(-\left(\hat{\hat{u}}_i - \min_{j=1,\ldots,n} \{\hat{\hat{u}}_i\}\right)\right), \quad i = 1, \ldots, n \tag{8}$$

The drawbacks of the COLS approach will now be discussed shortly. It is not clear whether it is only a philosophical weakness to get a shifted average function instead of a 'genuine' frontier (see e.g. Kalirajan and Obwona (1994)). Assuming the independence of inputs x and inefficiency u certainly can be problematic. But the most important difficulties – see section 5 – are due to the lack of an error term in the model. The deviations of the observations from the frontier are assumed to stem only from inefficiency.

2.2. *Stochastic frontiers*

Aigner et al. (1977) and Meeusen and van den Broek (1977) propose the stochastic frontier (SF) or composed error (CE) approach which has dominated both theory and applications since that time. They consider

$$y_i = \alpha + \sum_{j=1}^{k} \beta_j x_{ij} + e_i, \quad e_i = v_i - u_i, \quad u_i \geq 0, \quad i = 1, \ldots, n \tag{9}$$

where the composed error term e_i consists of a symmetric part v_i representing statistical noise and of the inefficiency term u_i following a one-sided distribution. It is assumed that v_i and u_i are independent. The distributional assumptions in section 4 will be

$$v_i \sim N(0, \sigma_v^2) \quad \text{and} \quad u_i \sim |N(0, \sigma_u^2)| \tag{10}$$

The log-likelihood function then is

$$l(\alpha, \beta, \sigma, \lambda) = -n\ln(\sigma) - const + \sum_{i=1}^{n}\left[\ln\Phi\left(\frac{-e_i\lambda}{\sigma}\right) - \frac{1}{2}\left(\frac{e_i}{\sigma}\right)^2\right] \tag{11}$$

with two 'variance parameters'

$$\lambda = \frac{\sigma_u}{\sigma_v} \quad \sigma^2 = \sigma_v^2 + \sigma_u^2 \tag{12}$$

and the moments

$$E(u_i) = \sqrt{\frac{2}{\pi}}\sigma_u \quad Var(u_i) = \left(\frac{\pi}{2} - 1\right)\sigma_u^2 \tag{13}$$

See Greene (1993) for the distributional alternatives for u.

The best way to estimate (9) is via iterative maximization of the log-likelihood (11), e.g. performed by LIMDEP (Greene (1989)). The ML estimators $\hat{\alpha}$ and $\hat{\beta}_j$ are consistent and asymptotically efficient.

The advantage of the stochastic frontier certainly is the inclusion of an error term. But there are disadvantages as well: In general, stochastic frontiers do not differ very much from a shifted average function (see Gong and Sickles (1992)). The independence of x and u has to be assumed like in the preceding COLS approach. But the greatest problem of the stochastic frontier is due to its advantage:

The estimation residuals estimate

$$e_i = y_i - \hat{\alpha} - \sum_{j=1}^{k}\hat{\beta}_j x_{ij}, \quad i = 1, \ldots, n \tag{14}$$

not u_i. Individual inefficiencies can only be estimated indirectly with the help of

$$u_i|e_i \sim trunc_0 N(\mu_i^*, \sigma_*^2), \quad \mu_i^* = \frac{-\sigma_u^2 e_i}{\sigma^2} \quad \sigma_* = \frac{\sigma_u\sigma_v}{\sigma} \tag{15}$$

by

$$E[u_i|e_i] = \frac{\sigma\lambda}{1+\lambda^2}\left(\frac{\phi(e_i\lambda/\sigma)}{\Phi(-e_i\lambda/\sigma)} - \frac{e_i\lambda}{\sigma}\right) \tag{16}$$

(Jondrow et al. (1982)). $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal density and cdf.

If the untransformed output $Y_i = \exp(y_i)$ is modeled instead of y_i, equation (9) becomes

$$Y_i = \exp(\alpha) \cdot \prod_{j=1}^{k}\exp(\beta_j x_{ij}) \cdot \exp(v_i) \cdot \exp(-u_i), \;\; u_i \geq 0, \;\; i = 1, \ldots, n \tag{17}$$

Then, since

$$E[e^{-u_i}|e_i] \neq e^{-E[u_i|e_i]} \tag{18}$$

appropriate point estimates for the untransformed individual technical efficiency $TE_i = \exp(-u_i)$ are

$$E[e^{-u_i}|e_i] = \exp\left(-\mu_i^* + \frac{1}{2}\sigma_*^2\right) \frac{1 - \Phi(\sigma_* - \mu_i^*/\sigma_*)}{1 - \Phi(-\mu_i^*/\sigma_*)} \tag{19}$$

(Battese and Coelli (1988)).

(16) and (19) provide unbiased but inconsistent estimators because the variance does not approach zero for large n. Consistency is only available with panel data where individual efficiency estimation is based on more than one observation per individual. It is simply impossible to seperate a sum of two unobservables into its components satisfactorily with one observation.

2.3. Résumé

As a result, the researcher has to select one out of three drawbacks. The standard OLS model lacks the inefficiency term u_i and consequently interprets inefficiency as misspecification. In the COLS approach, there is no disturbance term v_i. That is why measurement errors and stochastic variation are interpreted as inefficiency. The stochastic frontier comprises v_i and u_i but is no longer able to provide consistent estimates for u_i. A careful decision about this trade-off between model simplicity and descriptive accuracy depends on the individual economic problem (see Jensen (1999b)).

3. Example: Earnings frontier

The results in the following sections will be illustrated by an empirical example. Therefore, this section summarizes the necessary details of Jensen (1996 and 1999a) on the estimation of a stochastic earnings frontier.

Individual wages y are assumed to depend upon personal characteristics H augmenting human capital stock, job characteristics C and information I on labour market conditions, the wage distribution and job search methods. Individuals stop their search when a wage offer exceeds the reservation wage y_r. For any set of H and C and perfect information I^*, a potential maximum attainable wage y^* exists.

Then, $y = y(H, C, I)$ is estimated as stochastic earnings frontier (9) where y is empirical gross wage income in logs and $\hat{y}_i = y_i^*$ is estimated maximum possible income. x is a vector of $k = 22$ variables including a schooling dummy (qualification for university entrance), 2 dummies for professional training, 2 dummies for studies (technical college and university), experience (age and age squared), 2 dummies for on-the-job-training, 3 variables measuring ability, size of residence, dummies for sex and marital status, firm size, job status, dummies for employees and public servants, working time, seniority and a property variable. The inefficiency term u_i is interpreted as cost of

Table 1. SF: Ordered individual efficiencies

Efficiency rank	Efficiency
1	0.9612
2	0.9577
4	0.9465
8	0.9369
16	0.9274
32	0.9142
64	0.8856
128	0.8203
195	0.5461

imperfect information becoming apparent in the underemployment or over-education of individual number i.

The data set consists of $n = 1334$ individuals from the tenth wave (1992) of the German socio-economic panel (SOEP). See Jensen (1996 and 1999a) for more details on the data, the results and a critical assessment of the model. For the following sections, the complete sample and various subsamples have been analysed. The results presented in sections 4 and 5 are reasonably stable. That is why all figures and tables are prepared for a fixed subsample of size $n = 195$.

4. Efficiency ranking by stochastic frontiers

4.1. Efficiency ranking and confidence intervals

The stochastic earnings frontier of section 3 and equation (9) has been estimated and $n = 195$ individual inefficiency estimates $\widehat{e^{-u_i}}$ – not $e^{-\hat{u}_i}$ – have been calculated by (19). These estimates have been ordered from the most efficient to the most inefficient giving the efficiency ranking

$$\widehat{e^{-u_{[1]}}}, \widehat{e^{-u_{[2]}}}, \ldots, \widehat{e^{-u_{[195]}}} \tag{20}$$

Table 1 shows some selected ordered individual inefficiency estimates. A value of $e^{-u_i} = 1$ would mean no inefficiency ($u_i = 0$). According to the estimate, the most efficient individual in the sample receives

$$\widehat{e^{-u_{[1]}}} \cdot 100 \approx 96\% \tag{21}$$

of his/her maximum possible income.

Until Horrace and Schmidt (1996), in stochastic frontier models, almost nobody guarded these point estimates with confidence statements (see section 1 for the literature). Horrace and Schmidt derive the following two-sided confidence interval for e^{-u_i} from equation (15) and the monotonicity of the exponential function:

$$P[\exp(-\mu_i^* - z_L\sigma_*) \leq (e^{-u_i}|e_i) \leq \exp(-\mu_i^* - z_U\sigma_*)] = 1 - \gamma \tag{22}$$

Table 2. SF: Ordered individual efficiencies

Efficiency rank	Efficiency	95% confidence interval
1	0.9612	[0.8723, 0.9988]
2	0.9577	[0.8630, 0.9987]
4	0.9465	[0.8354, 0.9983]
8	0.9369	[0.8142, 0.9978]
16	0.9274	[0.7954, 0.9973]
32	0.9142	[0.7722, 0.9964]
64	0.8856	[0.7298, 0.9935]
▷ 75	0.8725	[0.7131, 0.9916]
128	0.8203	[0.6572, 0.9772]
▷ 149	0.7911	[0.6304, 0.9613]
195	0.5461	[0.4327, 0.6800]

where $Z \sim N(0,1)$ and

$$z_L = \Phi^{-1}\left\{1 - \frac{\gamma}{2}\left[1 - \Phi\left(-\frac{\mu_i^*}{\sigma_*}\right)\right]\right\} \tag{23}$$

$$z_U = \Phi^{-1}\left\{1 - \left(1 - \frac{\gamma}{2}\right)\left[1 - \Phi\left(-\frac{\mu_i^*}{\sigma_*}\right)\right]\right\} \tag{24}$$

See subsection 2.2 for the notation. Note that all parameters are treated as known which is unimportant only for large n.

Table 2 is table 1 supplied with an additional column of confidence intervals (and two additional rows of remarkable observations).

These confidence intervals overlap to a high degree. Horrace and Schmidt (1996) make the same observation and come to a pessimistic conclusion: '*… we found confidence intervals that were wider than we would have anticipated before this study began.*' (p. 276) '*However, frankly, in all cases that we considered the efficiency estimates were rather imprecise. We suspect that, in many empirical analyses using stochastic frontier models, differences across firms in efficiency levels are statistically insignificant, and much of what has been carefully explained by empirical analysts may be nothing more than sampling error.*' (p. 281).

Horrace and Schmidt examine efficiency differences. We will analyse the effects of confidence intervals on the reliability of efficiency rankings. And, in this connection, we will concentrate on the reliability of the selection of the most efficient individual. How sure can the individual on efficiency rank no. 1 be to be the best? But the results of the following pages can easily be applied to similar questions about rankings (being among the top 4, etc.).

Table 2 shows that the point estimates of efficiency ranks 1 to 75 are included in the confidence interval of efficiency rank 1. And the confidence intervals of efficiency ranks 1 to 149 include the point estimate of efficiency rank 1. This means that the pessimistic conclusion of Horrace and Schmidt (1996) applies to efficiency rankings as well (if one looks only at the individual confidence intervals).

Figure 1 presents the absolute frequencies of the upper bounds of the 95% intervals for $\exp(-u_i)$. Figure 2 shows the length of the confidence intervals

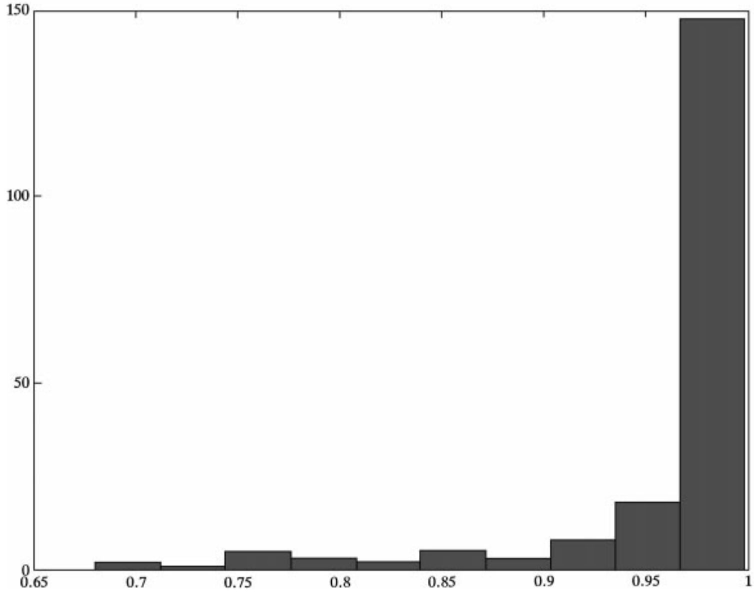

Fig. 1. SF: Abs. frequencies of upper bounds of 95% intervals for TE

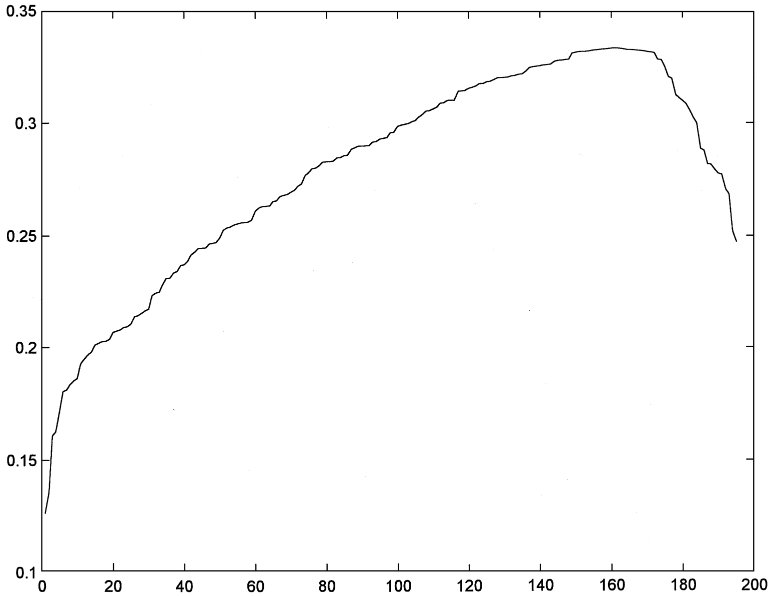

Fig. 2. SF: Length of confidence intervals as function of efficiency rank

as a function of the efficiency rank. The figures confirm what the last column of table 2 also tells: The confidence intervals overlap to such a high degree because the upper bounds don't get away fast enough from 0.99…

4.2. *Membership function*

Should we stop analysing individual efficiency estimates because this only means rummaging about in sampling error? Should we therefore even stop estimating frontier models because being able to provide reliable individual inefficiency estimates is a main feature of them? In this subsection, we will look for a way out of this dilemma.

Horrace and Schmidt (1996) find out that the confidence intervals become narrower when $\lambda = \sigma_u/\sigma_v$ is large and when – in the panel data case – T (the number of observations per individual) is large. They also work – in the panel data case – with the 'multiple comparisons with the best' technique (Edwards and Hsu (1983)) without enourmously reducing the interval length.

It has been seen in the previous subsection that the individual confidence intervals are not able to give much confidence in efficiency rankings. Therefore, we will now apply alternative ideas. One alternative would be to try to calculate the probability that individual no. i is the most efficient in the sample:

$$P\left[i = \arg\min_{j=1,\dots,n}(u_j)\right] = P[u_i \le u_j \mid j = 1,\dots,n] = P\left[u_i - \min_{j=1,\dots,n}(u_j) \le 0\right] \tag{25}$$

Of course, the calculation of this probability is possible but – because of the truncated normal distribution of $u_i|e_i$ in (15) – this is expected to involve complicated integrals.

Another idea would be to follow the 'subset selection approach' from the 'theory on statistical selection' (see e.g. Bechhofer et al. (1995) and the literature cited there) where the goal is to select a subset (as small as possible) of individuals that contains the most efficient individual with given probability. Once again, because of the truncated normal distribution there are no results available in the literature (to my knowledge) and easy solutions are not expected.

That is why we will pursue a very straightforward idea to concentrate the information in the sample on the most efficient individual. The following procedure provides simple Monte-Carlo approximations for the probability (25) and for subsets containing the most efficient individual.

If the individual on efficiency rank 149 should be the best in reality, his true efficiency value must be drastically better than his efficiency estimate and ALL other true efficiencies must be (in some cases drastically) worse than their estimates. This is very unlikely! Therefore, the proposed simulation procedure is: Simulate IT times (IT sufficiently large) a vector of realizations of

$$(e^{-u_1}|e_1,\dots,e^{-u_n}|e_n) \quad \text{with } u_i|e_i \sim trunc_0 N(\mu_i^*,\sigma_*^2) \quad \text{for } i = 1,\dots,n \tag{26}$$

(see equation (15)), identify the most efficient individual in each iteration and count the number $m(i)$ of iterations where individual i is the most efficient one. Then, define the membership function

$$\text{mf} : \{[1],[2],\dots,[n]\} \to [0,1], \quad [i] \mapsto \frac{m([i])}{IT} \tag{27}$$

Table 3. SF: Ordered individual efficiencies

Eff. rank	Eff.	95% conf. int.	memb. fct	cum. memb. fct.
1	0.9612	[0.8723, 0.9988]	0.1178	0.1178
2	0.9577	[0.8630, 0.9987]	0.0843	0.2020
4	0.9465	[0.8354, 0.9983]	0.0520	0.3088
8	0.9369	[0.8142, 0.9978]	0.0338	0.4268
16	0.9274	[0.7954, 0.9973]	0.0243	0.6308
▷ 30	0.9186	[0.7797, 0.9968]	0.0238	0.9503
32	0.9142	[0.7722, 0.9964]	0.0188	0.9875
▷ 34	0.9116	[0.7678, 0.9962]	0.0010	1
64	0.8856	[0.7298, 0.9935]	0	1
▷ 75	0.8725	[0.7131, 0.9916]	0	1
128	0.8203	[0.6572, 0.9772]	0	1
▷ 149	0.7911	[0.6304, 0.9613]	0	1
195	0.5461	[0.4327, 0.6800]	0	1

giving the relative frequencies of the maximum on the set of ordered individuals (ranked by the efficiency ranking).

Concerning the appropriate choice of the number of iterations IT, we have analysed $IT = 500$, 1000, 2000 and 4000. There are no substantial changes in the results but – of course – the membership function is smoothed with growing IT. So, the results for $IT = 4000$ are reported in the following.

Table 3 is table 2 supplied with two additional columns giving the membership function and the cumulative membership function

$$\mathrm{cmf}: \{[1], [2], \ldots, [n]\} \to [0, 1], \quad [i] \mapsto \sum_{j=1}^{i} \mathrm{mf}([j]) \tag{28}$$

(and two additional rows of remarkable observations). Figure 3 shows the membership function for the empirical example.

The membership function is different from zero only for the efficiency ranks 1 to 34 meaning that – if the Monte-Carlo approximation works properly – there is reason to suppose that one of these 34 individuals is the most efficient individual in the sample. The cumulative membership function crosses the value 0.95 for efficiency rank 30 meaning that the subset containing the ordered individuals 1 to 30 is probably the smallest possible subset containing the most efficient individual with probability 0.95. The membership function value for efficiency rank no. 1 means that with probability $p \approx 0.12$ this individual is the most efficient in the sample. Membership functions can easily be defined and calculated for similar questions about e.g. the set of individuals containing the top 4, etc.

It has been shown how the membership function can help to reduce uncertainty about the most efficient individual. But, of course, the properties of this approximation for probability calculation and subset selection still have to be explored. Furthermore, the figure suggests that some suitable smoothing could improve its accuracy. Finally, in small samples, the estimation error in $(\hat{\alpha}, \hat{\beta})$ might influence the results whereas the truncated normal distribution of $u_i|e_i$ in (15) only reflects the uncertainty about u_i due to v_i. A

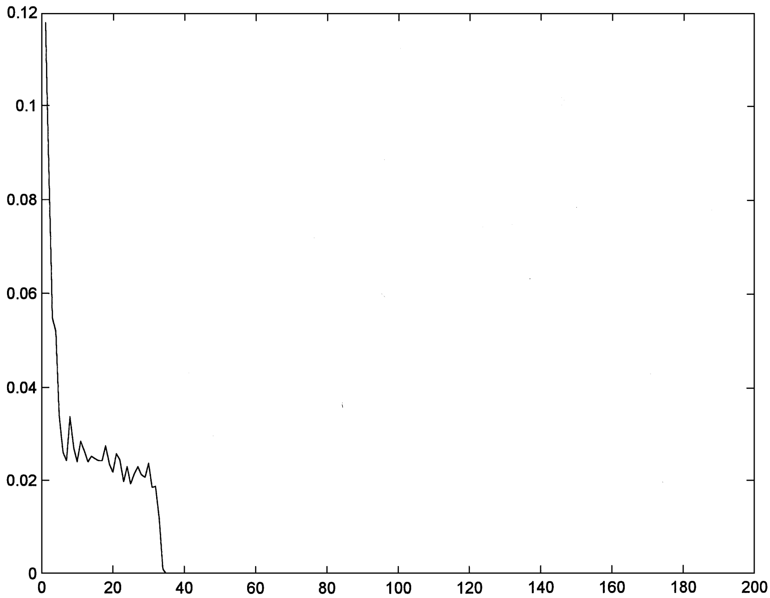

Fig. 3. SF: Membership function

bootstrap procedure reestimating the parameters might help in this respect. But the solution of these problems lies beyond the scope of this paper.

After all, even when approximating probabilites or subset selections with membership functions, uncertainty about the most efficient individual in the present sample cannot be reduced to less than being in a subset of roughly 15% individuals. So, this seems to be the price for choosing the stochastic frontier model (with inefficiency and noise) and having a cross-section dataset with only one observation per individual.

5. Efficiency ranking by COLS

5.1. *Efficiency ranking and bootstrap confidence intervals*

The stochastic frontier model is the only frontier model including an inefficiency term and a noise term. It is one of the main objections against the application of the non-parametric DEA or the deterministic LP or COLS approaches that they do not account for statistical noise – see the surveys in Greene (1993) and Lovell (1993). For comparison, in this section, we will therefore apply the COLS approach to the problem of identifying the most efficient individual in the sample.

The earnings frontier of section 3 has been estimated as COLS model (equations (3) and (5)). $n = 195$ individual inefficiency estimates $\widehat{TE}_i$ have been estimated following equation (8). Table 4 shows some selected values $e^{-\hat{u}_{[i]}}$ from the COLS efficiency ranking compared with the corresponding stochastic frontier efficiency ranks from table 1. The COLS results are

Table 4. Ordered individual efficiencies

Stoch. Front.		COLS	
Eff. rank	Efficiency	Eff. rank	Efficiency
1	0.9612	1	1.0000
2	0.9577	2	0.9704
4	0.9465	4	0.8614
8	0.9369	8	0.8038
16	0.9274	16	0.7578
32	0.9142	32	0.7170
64	0.8856	64	0.6475
128	0.8203	128	0.5598
195	0.5461	195	0.3336

interpreted like the stochastic frontier results in equation (21). But it has to be kept in mind that there is no error term in the model and, therefore, measurement errors and stochastic variation are interpreted as inefficiency. That is why the differences between the $e^{-\hat{u}_{[i]}}$ are larger (too large) than with the stochastic frontier.

The individual inefficiency differences estimated by COLS are too large. But, of course, this difference between COLS and SF does not carry over to the efficiency rankings. The approaches agree reasonably well about the most efficient and the most inefficient individuals, and efficiency ranks 1 to 4 are even identical. These findings are in line with earlier results on this question. And these findings are to be expected because stochastic frontiers are known to differ not very much from shifted OLS functions (see subsection 2.2) and because stochastic frontier rankings based on estimates for e_i, $E[u_i|e_i]$ or $E[e^{-u_i}|e_i]$ are identical – see equations (14), (16) and (19).

Because theoretically derived confidence intervals are not available for COLS residuals, we will now construct bootstrap confidence intervals for the e^{-u_i} – see Efron and Tibshirani (1993) or Shao and Tu (1995) for a general survey on the bootstrap and Hall et al. (1995) about bootstrap intervals for the intercept in a fixed effect panel frontier model. The proposed procedure employs the simple percentile method following an idea of Ferrier and Hirschberg (1997) for the construction of bootstrap confidence intervals for individual efficiencies in the DEA approach. The properties of this approach and possible alternatives will be discussed subsequently.

1. Estimate the vector of centered predicted residuals $\tilde{u}$ with

$$\hat{\hat{u}}_i = \hat{\hat{\alpha}} + \sum_{j=1}^{k} \hat{\beta}_j x_{ij} - y_i, \quad i = 1, \ldots, n \tag{29}$$

2. For every $i = 1, \ldots, n$:

   (a) For $b = 1, \ldots, B$ (B sufficiently large):

      i. Resampling: Denoting the bootstrap values of the bth resampling run with '$*b$', keep $\tilde{u}_i^{*b} := \hat{\hat{u}}_i$ fixed and obtain the remaining $n-1$

Table 5. COLS: Ordered individual efficiencies

Efficiency rank	Efficiency	95% confidence interval
1	1.0000	[0.9155, 1.0000]
$\triangleright$ 2	0.9704	[0.8671, 1.0000]
$\triangleright$ 4	0.8614	[0.7919, 1.0000]
8	0.8038	[0.7304, 0.9450]
16	0.7578	[0.6990, 0.8866]
32	0.7170	[0.6386, 0.8519]
64	0.6475	[0.6028, 0.7622]
128	0.5598	[0.5171, 0.6677]
195	0.3336	[0.3167, 0.4312]

residuals

$$\tilde{u}_1^{*b}, \dots, \tilde{u}_{i-1}^{*b}, \tilde{u}_{i+1}^{*b}, \dots, \tilde{u}_n^{*b} \tag{30}$$

by drawing $n-1$ times with replacement from $\hat{\tilde{u}}$.

ii. Construct the 'fake data' y^{*b} by

$$y_l^{*b} = \hat{\tilde{\alpha}} + \sum_{j=1}^{k} \hat{\beta}_j x_{lj} - \tilde{u}_l^{*b}, \quad l = 1, \dots, n \tag{31}$$

iii. Using x and y^{*b}, calculate the reestimates $(\widehat{\alpha^{*b}}, \widehat{\beta^{*b}}, \widehat{\tilde{u}^{*b}})$ and

$$\widehat{TE_i^{*b}} = \exp(-\widehat{u_i^{*b}}) = \exp\left(-\left(\widehat{\tilde{u}_i^{*b}} - \min_{j=1,\dots,n}(\widehat{\tilde{u}_j^{*b}})\right)\right) \tag{32}$$

(b) The $(1-2\gamma)$ confidence interval for TE_i is the interval $[\widehat{TE_i^{*\gamma}}, \widehat{TE_i^{*(1-\gamma)}}]$ where $\widehat{TE_i^{*\gamma}}$ and $\widehat{TE_i^{*(1-\gamma)}}$ are the 100γ and $100(1-\gamma)$ percentiles of the bootstrap distribution of $\widehat{TE_i^{*}}$.

$B = 1000$ and $B = 2000$ resampling runs have been analysed, and the vast majority of confidence intervals becomes longer when increasing the number of resampling runs from $B = 1000$ to $B = 2000$. Because of the discussion around equation (33), the results for $B = 2000$ are reported in the following. Note that $n \cdot B = 390,000$ resampling runs are needed for the construction of $n = 195$ confidence intervals. The result of every resampling run is one single $\widehat{TE_i^{*b}}$ giving a $(n \times B)$ matrix $\widehat{TE^{*}}$ we will apply in the following subsection. Now, table 5 is the right-hand side of table 4 with an additional column of 95% bootstrap confidence intervals.

It can be seen that – in contrast to the results of table 2 – only the point estimates of efficiency ranks 1 and 2 are included in the confidence interval of efficiency rank 1. And only the confidence intervals of efficiency ranks 1 to 4 include the point estimate of efficiency rank 1. Figure 4 shows the absolute frequencies of the upper bounds of the 95% intervals for $TE_i = \exp(-u_i)$. Figure 5 presents the length of the bootstrap confidence intervals as a function of the efficiency rank. The message of the figures is in line with the findings in

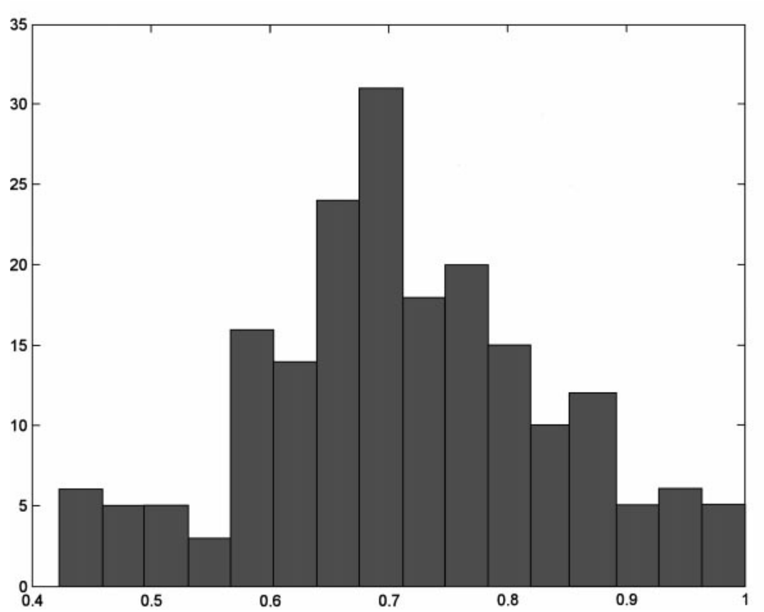

Fig. 4. COLS: Abs. frequencies of upper bounds of 95% intervals for TE

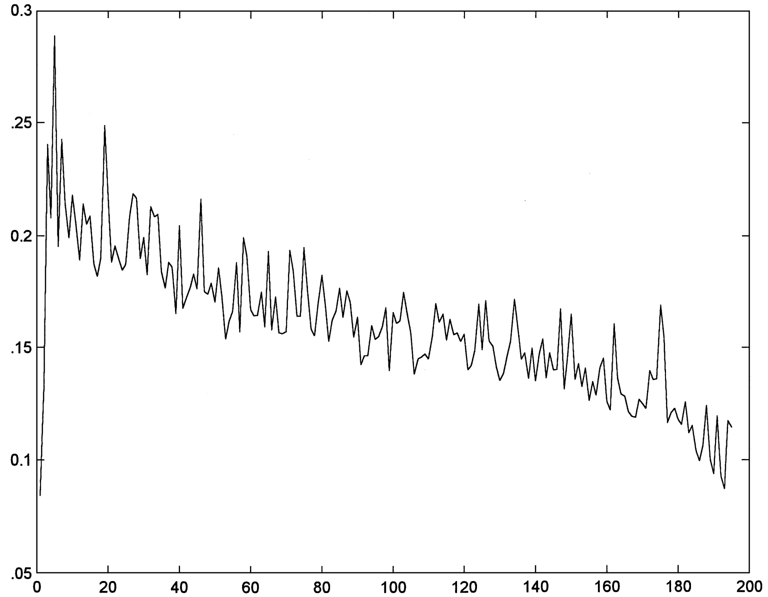

Fig. 5. COLS: Length of bootstrap intervals as function of efficiency ranks

the table because the upper bounds do not stick to 1 (and most of the intervals are shorter as in section 4.1).

Comparing theses results with table 2 and figures 1 and 2 leads to the conclusion that COLS (in connection with bootstrap confidence intervals) is very 'powerful' in its decision about the most efficient individual. Why is this the case? Is this power reliable?

Two reasons can be given for the fact that the COLS bootstrap intervals do not overlap as much as the stochastic frontier intervals. First, as mentioned at the beginning of this section, the differences between the TE_i or u_i are too large because there is no error term in the model and, therefore, measurement errors and stochastic variation are included in the inefficiency term.

Secondly, the bootstrap intervals mostly are too short. The simple percentile method is known to have problems in the tails when the theoretical distribution has an infinite support whereas the sample support is finite (see e.g. Efron and Tibshirani (1993)). Of course, this effect gets worse with smaller sample sizes. From the theoretical side, the simple percentile method is only first-order accurate meaning

$$P[TE_i \leq \widehat{TE_i^{*\gamma}}] = \gamma + O(n^{-1/2}) \tag{33}$$

whereas improved techniques like BC_α, percentile-t or the iterated percentile method often are second-order accurate (see e.g. Shao and Tu (1995) for detailed results).

Unfortunately, the use of the percentile-t and BC_α is limited because the former requires a good variance estimator whereas the latter requires a good estimator of the 'acceleration parameter' (see Shao and Tu (1995, p. 151)). Therefore, the iterated percentile method is recommended in Jeong and Maddala (1993) and Hall et al. (1995). But this method is '... *almost prohibitively laborious* ...' (Hall (1986)). If C denotes the number of re-resampling runs, it would take $n \cdot B \cdot C \approx 195{,}000{,}000$ (re-)resampling runs to construct $n = 195$ confidence intervals (see Booth and Presnell (1998) for appropriate choices of B and C). That is why the simple percentile method has been chosen despite its drawbacks.

But Hall et al. (1993 and 1995) show even another potential problem of the simple percentile method in this application. If the goal is to construct a bootstrap confidence interval for the ith largest $TE_{[i]}$, the bootstrap estimator of the distribution of $\widehat{TE}_{[i]}$ is consistent if and only if there are no ties for $TE_{[i]}$. Of course, exact ties are very unlikely. But, in small to moderate samples, two or more close values competing for $TE_{[i]}$ can cause coverage inaccuracies of the bootstrap confidence interval for $TE_{[i]}$. Therefore, Hall et al. (1995) suggest to correct for coverage error with the iterated bootstrap.

Because of these three disadvantages, the (simple percentile) bootstrap intervals in the COLS approach have to be interpreted very cautiously. But, in the next subsection, the membership function will at least help avoiding the second problem of this approach.

Finally, some comments are due to the similarities and differences between the bootstrap procedure in the COLS approach in this paper and the bootstrap procedure for the DEA approach by Ferrier and Hirschberg (1997). Both procedures are developed for the construction of bootstrap confidence intervals for individual efficiencies, and this paper has adopted the idea of Ferrier and Hirschberg of keeping the i-th centered predicted residual $\tilde{u}_i^{*b} := \hat{\hat{u}}_i$ fixed in all $b = 1, \ldots, B$ resampling runs (see step 2.(a)i. of the bootstrap procedure) when constructing the confidence interval for the i-th individual efficiency. But apart from that, there are no important similarities of the two bootstrap procedures.

Table 6. COLS: Ordered individual efficiencies

Eff. rank	Eff.	95% conf. int.	memb. fct.	cum. memb. fct.
1	1.0000	[0.9155, 1.0000]	0.3705	0.3705
▷ 2	0.9704	[0.8671, 1.0000]	0.2460	0.6165
▷ 4	0.8614	[0.7919, 1.0000]	0.0820	0.8890
▷ 7	0.8069	[0.7050, 0.9475]	0.0220	0.9585
8	0.8038	[0.7304, 0.9450]	0.0165	0.9750
▷ 14	0.7687	[0.6961, 0.9015]	0.0005	1
16	0.7578	[0.6990, 0.8866]	0	1
32	0.7170	[0.6386, 0.8519]	0	1
64	0.6475	[0.6028, 0.7622]	0	1
128	0.5598	[0.5171, 0.6677]	0	1
195	0.3336	[0.3167, 0.4312]	0	1

This is essential since the bootstrap procedure by Ferrier and Hirschberg (1997) has been heavily criticized in a recent paper by Simar and Wilson (1999) for being inconsistent. This inconsistency of the simple percentile method in the DEA approach occurs for efficient observations because there can be probability mass at the frontier in the bootstrap distribution. Table 2 of Ferrier and Hirschberg (1997, p. 26f) shows that 28 of the 94 bootstrap distributions for the TE_i are even degenerate (mean 1 and variance 0). But this problem is a speciality of the non-parametric DEA estimator and does not carry over (see table 5) to the COLS approach based on the linear model (because there is variation in the regressors).

One referee has asked why the i-th centered predicted residual $\tilde{u}_i^{*b}$ is fixed in all resampling runs in step 2.(a)i. of the bootstrap procedure. The reason is that I don't see any better way of calculating confidence intervals especially for the efficiency of a very efficient (or very inefficient) individual i. These individuals ARE very (in)efficient and, consequently, this property has to enter in the resampling runs. But since $\tilde{u}_j^{*b}$ for $j \neq i$ is drawn from the complete vector $\hat{\hat{u}}$, many other individuals have the opportunity to be more efficient than i in some resampling runs.

In any case, keeping the i-th centered predicted residual fixed does not overcome the inconsistency of the Ferrier and Hirschberg procedure (see Simar and Wilson (1999, p. 73ff)) and it does not cause inconsistency in the bootstrap procedure of this paper.

5.2. Membership function

In the stochastic frontier case, the calculation of the membership function required a laborious simulation procedure. In case of COLS, we simply apply the $(n \times B)$ bootstrap matrix $\widehat{TE}^*$: Count the number of columns of $\widehat{TE}^*$ where individual i is the most efficient one and calculate the membership function and the cumulative membership function according to the definitions (27) and (28). All comments in subsection 4.2 concerning properties, smoothing and interpretation apply as well.

Table 6 is table 5 supplied with two additional columns giving the membership function and the cumulative membership function (and two additional

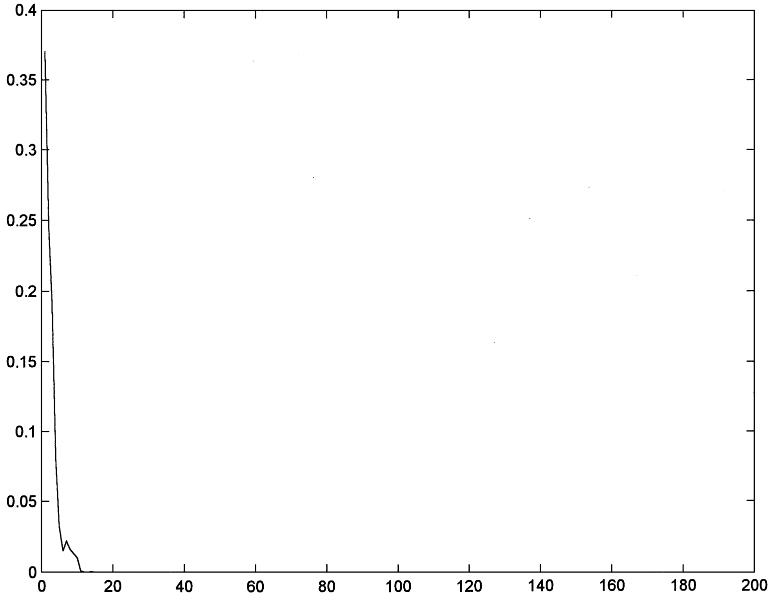

Fig. 6. COLS: Membership function

rows with remarkable observations). Figure 6 presents the membership function for the empirical example. The membership function is different from zero for the efficiency ranks 1 to 14, and the cumulative membership function crosses the value 0.95 for individual no. 7.

Whereas the membership function reduces the uncertainty about the most efficient individual in the stochastic frontier approach, it increases the uncertainty in this case. This had to be expected because now the decision about the most efficient individual is made directly with the $\widehat{TE}^*$ matrix and not with the simple percentile bootstrap intervals which are known to be too short.

6. Conclusions

In this paper, we have analysed the consequences of the consideration of confidence statements for the reliability of efficiency rankings. In this connection, we have concentrated on the reliability of the selection of the most efficient individual. Regarding only confidence intervals, the stochastic frontier approach is very powerless confirming the disappointed result of Horrace and Schmidt (1996): Efficiency differences across individuals or efficiency rankings of individuals are probably often 'nothing more than sampling error'. In the cross-section case, the inconsistency of individual efficiency estimates (one observation per individual) once more turns out to be problematic. But with the aid of the membership function, this uncertainty could be reduced to some extent.

Alternatively, the COLS ranking was analysed and a procedure for the calculation of bootstrap confidence intervals for the individual inefficiency estimates was proposed. This approach seems to be very powerful in the decision on the most efficient individual. But Cornwell and Schmidt (1996) are right to remark that this precision is only faked: '*Using a deterministic model does not remove this uncertainty; it only hides it.*' This warning addresses not only to COLS, but also to the LP approach and the very popular DEA. Because there are no error terms in these models, estimated individual inefficiency differences are differences in inefficiency and error. And only bootstrapping a badly specified 'residual' does NOT provide a stochastic model with inefficiency term and noise. Secondly, the confidence intervals turned out to be too small because the simple percentile method is known to have problems in the tails. This problem could be by-passed by the application of the membership function. But finally, higher coverage inaccuracy can occur because of 'near-inconsistency' of the bootstrap procedure.

Summing up, it may be said that it IS efficient to analyse efficiency rankings if you take into account the inefficiency of the techniques. Confidence statements should always be mentioned. Suppressing them does not increase the precision of the results. At least if one is interested in individual efficiency estimation, frontier models should allow for inefficiency and noise. Suppressing the latter leads to overestimation of precision.

Finally, analysing cross-section data (and not panel data with large T) seems to lead very often to very imprecise individual efficiency estimates. That is why only large inefficiency differences or large efficiency rank differences should be regarded as serious. In an evaluation of the research and teaching quality of 100 universities, a university on efficiency rank 15 might as well be on the first rank. And analysing this efficiency (rank) difference may be 'much ado about nothing more than sampling error'.

References

Aigner DJ, Chu SF (1968) On estimating the industry production function. American Economic Review 58:826–839

Aigner DJ, Lovell CAK, Schmidt P (1977) Formulation and estimation of stochastic frontier production function models. Journal of Econometrics 6:21–37

Backes-Gellner U (1989) Ökonomie der Hochschulforschung. Gabler, Wiesbaden

Battese GE, Coelli TJ (1988) Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. Journal of Econometrics 38:387–399

Bechhofer RE, Santner TJ, Goldsman DM (1995) Design and analysis of experiments for statistical selection, screening, and multiple comparisons. Wiley, New York

Booth J, Presnell B (1998) Allocation of Monte Carlo resources for the iterated bootstrap. Journal of Computational and Graphical Statistics 7:92–112

Caudill SB, Ford JM, Gropper DM (1995) Frontier estimation and firm-specific inefficiency measures in the presence of heteroscedasticity. Journal of Business & Economic Statistics 13:105–111

Cornwell C, Schmidt P (1996) Production frontiers and efficiency measurement. In: Mátyás L, Sevestre P (ed) Econometrics of panel data: Handbook of theory and applications, 2. ed., Kluwer, Boston, pp. 845–878

Edwards DG, Hsu JC (1983) Multiple comparisons with the best treatment. Journal of the American Statistical Association 78:965–971

Efron B, Tibshirani RJ (1993) An introduction to the bootstrap. Chapman & Hall, New York

Farrell MJ (1957) The measurement of productive efficiency. J.R.S.S. A 120:253–290

Ferrier GD, Hirschberg JG (1997) Bootstrapping confidence intervals for linear programming efficiency scores: With an illustration using Italian banking data. Journal of Productivity Analysis 8:19–33

Førsund FR, Lovell CAK, Schmidt P (1980) A survey of frontier production functions and of their relationship to efficiency measurement. Journal of Econometrics 13:5–25
Gong BH, Sickles RC (1992) Finite sample evidence on the performance of stochastic frontiers and data envelopment analysis using panel data. Journal of Econometrics 51:259–284
Greene WH (1980) Maximum likelihood estimation of econometric frontier functions. Journal of Econometrics 13:27–56
Greene WH (1989) LIMDEP. Econometric Software Inc., New York
Greene WH (1993) The econometric approach to efficiency analysis. In: Fried HO, Lovell CAK, Schmidt SS (ed) The measurement of productive efficiency, Oxford University Press, New York, pp. 68–119
Hall P (1986) On the bootstrap and confidence intervals. Annals of Statistics 14:1431–1452
Hall P, Härdle W, Simar L (1993) On the inconsistency of bootstrap distribution estimators. Computational Statistics & Data Analysis 16:11–18
Hall P, Härdle W, Simar L (1995) Iterated bootstrap with applications to frontier models. Journal of Productivity Analysis 6:63–76
Horrace WC, Schmidt P (1996) Confidence statements for efficiency estimates from stochastic frontier models. Journal of Productivity Analysis 7:257–282
Jensen U (1996) Estimation of an earnings frontier with data from the German socio-economic panel. Arbeiten aus dem Institut für Statistik und Ökonometrie, Nr. 92, Universität Kiel
Jensen U (1999a) Measuring earnings differentials with frontier functions and Rao distances. In: Marriott P, Salmon M (ed) Applications of differential geometry to econometrics, Cambridge University Press, to appear
Jensen U (1999b) The simplicity of an earnings frontier. In: Keuzenkamp HA, McAleer M, Zellner A (ed) Simplicity, Cambridge University Press, to appear
Jeong J, Maddala GS (1993) A perspective on application of bootstrap methods in econometrics. In: Maddala GS, Rao CR, Vinod HD (ed) Handbook of statistics, Vol. 11, Elsevier, pp. 573–610
Jondrow J, Lovell CAK, Materov IS, Schmidt P (1982) On the estimation of technical inefficiency in the stochastic frontier production function model. Journal of Econometrics 19:233–238
Kalirajan KP, Obwona MB (1994) Frontier production function: The stochastic coefficients approach. Oxford Bulletin of Economics and Statistics 56:87–96
Lovell CAK (1993) Production frontiers and productive efficiency. In: Fried HO, Lovell CAK, Schmidt SS (ed) The measurement of productive efficiency, Oxford University Press, New York, pp. 3–67
Meeusen W, van den Broeck J (1977) Efficiency estimation from Cobb-Douglas production functions with composed error. International Economic Review 18:435–444
Reifschneider D, Stevenson R (1991) Systematic departures from the frontier: A framework for the analysis of firm inefficiency. International Economic Review 32:715–723
Schmidt P (1976) On the statistical estimation of parametric frontier production functions. Review of Economics and Statistics 58:238–239
Shao J, Tu D (1995) The jackknife and bootstrap. Springer, New York
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
Simar L, Wilson PW (1998) Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. Management Science 44:49–61
Simar L, Wilson PW (1999) Some problems with the Ferrier/Hirschberg bootstrap idea. Journal of Productivity Analysis 11:67–80
Taube R (1988) Möglichkeiten der Effizienzmessung von öffentlichen Verwaltungen. Duncker & Humblot, Berlin
Winsten CB (1957) Discussion on Mr. Farrell's paper. J.R.S.S. A 120:282–284