ELSEVIER
Available online at www.sciencedirect.com
SCIENCE DIRECT®
European Journal of Operational Research xxx (2005) xxx–xxx
EUROPEAN JOURNAL OF OPERATIONAL RESEARCH
www.elsevier.com/locate/ejor

Interfaces with Other Disciplines

The super-efficiency procedure for outlier identification, not for ranking efficient units

Rajiv D. Banker [a], Hsihui Chang [b,*]

[a] *Fox School of Business Temple University Philadelphia, PA 19122, USA*
[b] *Anderson Graduate School of Management University of California, Riverside, CA 92521, USA*

Received 3 December 2004; accepted 28 June 2005

Abstract

In this paper, we conduct simulation experiments to evaluate the performance of two alternative uses of the super-efficiency procedure in Data Envelopment Analysis (DEA). The first is for outlier identification and the second is for ranking efficient units. We find that the ranking procedure does not perform satisfactorily. In fact, the correlations between the true efficiency and the estimated super-efficiency are negative for the subset of efficient observations, and the conventional DEA model performs as well as the super-efficiency DEA model when all observations are considered. However, when data are contaminated with outliers, the use of the super-efficiency model to identify and remove outliers results in more accurate efficiency estimates than those obtained from the conventional DEA estimation model.

Keywords: DEA; Super-efficiency; Outlier identification; Efficiency ranking; Simulation study

1. Introduction

Conventional models in data envelopment analysis (DEA) evaluate the efficiency of an observation relative to a reference set comprising of all sample observations, including itself. In contrast, the super-efficiency model in DEA excludes each observation from its own reference set, so that it is possible to obtain efficiency scores that exceed one. Banker and Gifford (1988) suggested the use of the super-efficiency model to screen out observations with gross data errors, and obtain more reliable efficiency estimates after removing those identified outliers. Banker et al. (1989) applied this method for outlier identification to analyze cost variances for 117 hospitals. Andersen and Petersen (1993) employed the same Banker–Gifford model and prescribed the use of the super-efficiency score

* Corresponding author. Tel.: +1 951 827 4284.
E-mail addresses: banker@temple.edu (R.D. Banker), hsihui.chang@ucr.edu (H. Chang).

0377-2217/$ - see front matter © 2005 Published by Elsevier B.V.
doi:10.1016/j.ejor.2005.06.028

for ranking efficient units in DEA. In this paper, we conduct simulation experiments to evaluate the performance of these two procedures. We find that the Andersen–Petersen procedure does not perform satisfactorily in ranking efficient units. We also find that the Banker–Gifford method for identifying and removing outliers in DEA performs well when some observations are contaminated.

The Banker–Gifford (BG hereafter) method is similar in spirit to Timmer's (1971) procedure of peeling off a fraction of the observations to obtain more reliable production frontier estimates. This procedure is designed for situations when some observations may be contaminated and, consequently, erroneously classified as efficient. The first step in the BG method identifies as outliers those observations whose super-efficiency score exceeds a pre-specified screen level (see Banker et al., 1989, pp. 279–280).[1] In the second step, the observations identified as outliers are removed, and a conventional DEA model, such as the BCC model (Banker et al., 1984) is estimated with the remaining observations.

The BG model was also employed to extend sensitivity analysis of DEA efficiency scores to proportionate changes in inputs or outputs (see Charnes and Neralic, 1990, 1992). This line of inquiry was followed by Andersen and Petersen (AP hereafter) who proposed a very different use for the BG model and referred to it as the super-efficiency DEA model. They suggested that the super-efficiency score be used to rank units that are found to be efficient using a conventional DEA model. Although all the efficient units have the same conventional efficiency score of one, their super-efficiency scores (greater than one) may be different. This provides the motivation for discriminating between efficient units using the AP procedure.

The objective of this paper is to evaluate the performance of these two alternative uses of the Banker–Gifford super-efficiency model based on simulation experiments. We evaluate how well the AP procedure performs in ranking efficient units when simulated data are not contaminated. We also evaluate the performance of the BG procedure for identifying and removing outliers when a small percentage of simulated data are contaminated with random noise in varying amounts.

The remainder of this paper is structured as follows. Section 2 describes the super-efficiency model in DEA. Section 3 describes the data generating process for the simulation experiments to evaluate the performance of the super-efficiency score in ranking efficient units when data are not contaminated. Section 4 describes the data generating process and simulation results to evaluate the performance of the super-efficiency model in outlier identification when some observations are contaminated with random noise. Section 5 concludes with a summary of our principal results.

2. Super-efficiency model

Let $Y_j \geqslant 0$ and $X_j \geqslant 0$, $j = 1, \ldots, N$, be the output and input vectors for N observations, with at least one element of each vector being strictly positive. The output-oriented super-efficiency measure $\hat{\psi}_k^{\mathrm{SE}}$ for an observation (X_k, Y_k), $k \in \{1, \ldots, N\}$ is the reciprocal of the super-inefficiency measure $\hat{\theta}_k^{\mathrm{SI}}$ obtained by solving the following linear program:

$$\hat{\theta}_k^{\mathrm{SI}} = \mathrm{Max}\,\theta_k \tag{1}$$

subject to

$$\sum_{\substack{j=1 \\ j \neq k}}^{N}, \lambda_j Y_j - \theta_k Y_k \geqslant 0, \tag{1a}$$

$$\sum_{\substack{j=1 \\ j \neq k}}^{N} \lambda_j X_j \leqslant X_k, \tag{1b}$$

$$\sum_{\substack{j=1 \\ j \neq k}}^{N} \lambda_j = 1, \tag{1c}$$

$$\theta_k, \quad \lambda_j \geqslant 0. \tag{1d}$$

Note that the difference between the above super-efficiency model and the conventional BCC model

[1] Subsequently, Wilson (1995) also suggested the same use of the super-efficiency model for identifying outliers.

is that when the super-efficiency model is employed, the observation "k" under evaluation is not included in the reference set for the constraints in (1a–1c). That is, the reference observation $\left(\sum_{j=1, j\neq k}^{N}\lambda_j X_j, \sum_{j=1, j\neq k}^{N}\lambda_j Y_j\right)$ in the evaluation of the super-efficiency of observation k is constructed only from observations other than k itself.

Since the observation k under evaluation is excluded from the reference set in the super-efficiency model, we cannot be sure that a convex combination can be created from the remaining observations to envelop observation k from under for its inputs and from above for its outputs. Banker and Gifford (1988) proved that while there always exists a feasible solution to the super-efficiency model for the CCR specification, there may not be a feasible solution to the super-efficiency model for the BCC specification for certain extreme observations.[2]

To avoid the computational problem associated with infeasible programs for the BCC super-efficiency model, we solve the following modified model:

$$\text{Max}\, \eta_k - 2\lambda_k \tag{2}$$

subject to

$$\sum_{j=1}^{N}\lambda_j Y_j - \eta_k Y_k \geqslant 0, \tag{2a}$$

$$\sum_{j=1}^{N}\lambda_j X_j \leqslant X_k, \tag{2b}$$

$$\sum_{j=1}^{N}\lambda_j = 1, \tag{2c}$$

$$\eta_k, \quad \lambda_j \geqslant 0. \tag{2d}$$

Because of the large negative weight on λ_k in the objective function in (2), an observation k will not serve as a reference point for its own evaluation (i.e., $\lambda_k^* = 1$) unless the corresponding problem in (1) is infeasible. Therefore, the super-efficiency of observation k is $\hat{\psi}_k^{\text{SE}} = 1/\eta_k^*$ if $\lambda_k^* = 0$ in an optimal solution to (2), and is marked as infeasible if $\lambda_k^* = 1$.[3] Observe that an observation k would be rated as inefficient by the conventional BCC model (that allows an observation to be in the reference set for itself, rather than excluding it as in program (1)), if and only if the super-efficiency estimate $\hat{\psi}_k^{\text{SE}} < 1$, and that the observation would have been rated as efficient by the conventional BCC model if and only if $\hat{\psi}_k^{\text{SE}} \geqslant 1$ (Banker and Gifford, 1988).

3. Ranking efficient units

As mentioned earlier, Andersen and Petersen (AP) suggested the use of the super-efficiency model for ranking efficient units. To evaluate the performance of the AP procedure in ranking efficient units, we conducted 1000 simulation experiments described below.

3.1. Data generating process

We considered three factors: sample size, production technology, and inefficiency distribution in generating the data for the simulation experiments reported in this section.

3.1.1. Sample size

For each experiment, we considered a sample of size N, where N can take any integer value between 51 and 150 with equal probability.

3.1.2. Production technology

We considered a single output and two inputs production technology specified in terms of its effi-

[2] While the super-efficiency model is specified in Eq. (1) for the BCC model, similar models can be specified corresponding to other DEA specifications. The CCR (Charnes et al., 1978) super-efficiency specification is the same linear program as in (1) except that the constraint in (1c) is deleted. The IRS (increasing returns to scale) specification is the same linear program as in (1) except that the constraint in (1c) is a less-than-or-equal-to ($\leqslant$) constraint instead of an equality ($=$) constraint. The DRS (decreasing returns to scale) specification is the same linear program as in (1) except that the constraint in (1c) is a greater-than-or-equal-to ($\geqslant$) constraint instead of an equality ($=$) constraint. Seiford and Zhu (1999) further extend the BG infeasibility result by proving the feasibility of the super-efficiency model for the IRS specification. See also Chen (2000) for combining output-oriented and input-oriented approaches to super-efficiency measurement.

[3] Observe that there always exists an optimal solution to (2) with either $\lambda_k^* = 0$ or $\lambda_k^* = 1$.

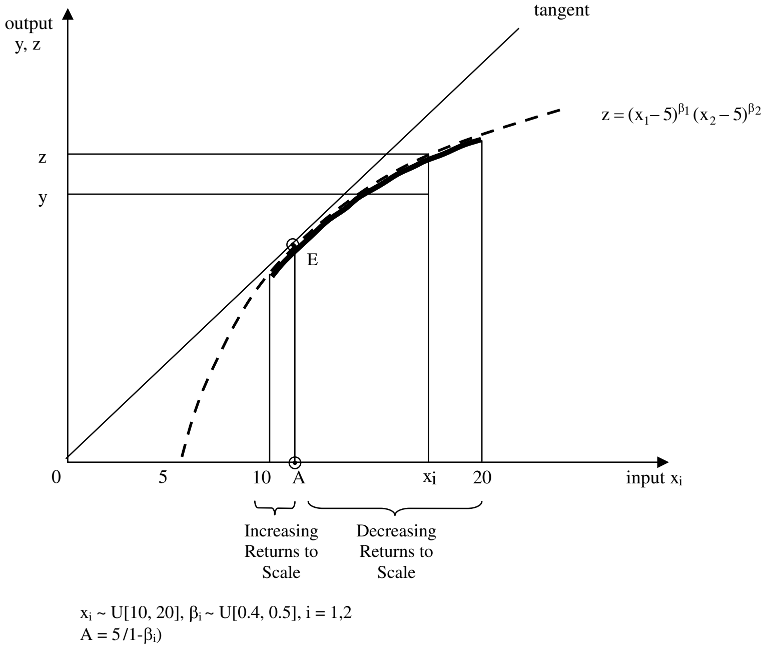

Fig. 1. Production function for the simulation experiments (relation between output z and input x_i for a fixed value of the other input x_{3-i}).

cient production function $z = f(x_1,\ x_2)$, where z represents the maximum output that can be produced from the levels x_1 and x_2 of the two inputs. Specifically, we used the following "shifted"[4] Cobb–Douglas production function

$$z = (x_1 - \alpha_1)^{\beta_1}(x_2 - \alpha_2)^{\beta_2}, \tag{3}$$

where $\alpha_1 = \alpha_2 = 5$. The inputs x_1 and x_2 were generated randomly from independent uniform probability distributions over the interval $[10, 20]$, and the coefficients β_1 and β_2 were generated randomly from independent uniform probability distributions over the interval $[0.4, 0.5]$. Since the sum of β_1 and β_2 is less than one, the production function in (3) satisfies the BCC model's maintained assumption of a concave production function, while the shifts $\alpha_1,\ \alpha_2 > 0$ allow both increasing and decreasing returns to scale to prevail. See Fig. 1 which portrays increasing returns over the interval $[10, A]$ and decreasing returns over the interval $[A, 20]$, where $A = 5/(1 - \beta_i)$, for the input x_i in a section of the production function obtained by fixing the level of the other input x_{3-i}, $i = 1, 2$.

3.1.3. Inefficiency distribution

We generated the logarithm of the inefficiency $u_k = \ln\theta_k$ for each observation $k \in \{1, \ldots, N\}$ from a half-normal distribution $|N(0, \sigma_u^2)|$, where the parameter σ_u^2 itself is drawn from a uniform distribution on the interval $[0, 0.1989]$. The range of values for the distribution of σ_u^2 is chosen such that mean efficiency given by $E(\psi) = E(e^{-u}) = \exp(-\sigma_u\sqrt{2/\pi})$ is between 0.7 and 1.0.

3.1.4. Simulated observations

For each experiment, we first randomly generated a value for N between 51 and 150, values for β_1 and β_2 between 0.4 and 0.5, and a value for σ_u^2 between 0 and 0.1989. Next, we simulated N observations of the two inputs x_1 and x_2 between 10 and 20. These values (x_{1k}, x_{2k}), $k = 1, \ldots, N$, were then substituted into the effi-

[4] We chose the shifted Cobb–Douglas because it provides a parsimonious parametric form that exhibits both increasing and decreasing returns to scale for $\alpha_i > 0$.

cient production function specified in (3) to obtain the corresponding values $z_k = f(x_{1k}, x_{2k})$ for the efficient output quantity. Then, we randomly generated the logarithm of "true" inefficiency values $u_k = \ln\theta_k$ for each observation $k \in \{1, \ldots, N\}$ from the half-normal distribution $|N(0, \sigma_u^2)|$. Finally, we obtained the values for "observed" output quantities y_k and the "true" efficiency values ψ_k as

$$y_k = f(x_{1k}, x_{2k})/\exp(u_k), \quad (4)$$

$$\psi_k = 1/\exp(u_k). \quad (5)$$

Thus, each observation k comprises its "observed" output and inputs values $(y_k; x_{1k}, x_{2k})$ and each sample consists of N such observations.

3.2. Simulation results

For each experiment, we first ran three linear programs in order to estimate the BCC efficiency and the BCC super-efficiency $\hat{\psi}_k^{\mathrm{SE}}$ for each of the $k = 1, \ldots, N$ observations in the sample.[5]

In Table 1 we report the means and medians for both Pearson and Spearman correlation coefficients between the true efficiency value ψ_k and the feasible super-efficiency estimates $\hat{\psi}_k^{\mathrm{SE}}$ for those observations rated as efficient by the BCC model. It is evident from Table 1 that the AP procedure's performance in ranking efficient units is not at all satisfactory. Both mean and median correlation coefficients are negative. In fact, as many as 706 (667) out of the 1000 experiments have a negative Pearson (Spearman) correlation coefficient between the true and super-efficiencies.

To examine the differences in the AP procedure's performance across different experiments, we regressed the Pearson correlation coefficients on three variables: the sample size N, the mean inefficiency $E(e^u) = \exp(\sigma_u\sqrt{2/\pi})$, and the production function concavity $(\beta_1 + \beta_2)$. The regression results reported in Table 2 reveal low R^2. While the degree of concavity of the production function and the mean inefficiency do not have a significant impact on the performance of the AP procedure using super-efficiency estimates, the performance deteriorates significantly when the sample size is larger. Further, regression results (not reported here) after including an additional variable $\delta = \beta_1 - \beta_2$ or $|\delta| = |\beta_1 - \beta_2|$ are very similar to those reported in Table 2 and the corresponding coefficient estimate for δ or $|\delta|$ is not significant.

Finally, for comparison purposes, we report in Table 3, mean and median correlation coefficients between the true and BCC and true and super-efficiency scores, where the correlation coefficients are calculated over all sample observations and not just those rated as efficient. Interestingly, we observe that the performance of the BCC procedure is quite comparable to that of the super-efficiency procedure.

Table 1
Means and medians of correlation coefficients between the true and super-efficiencies when data are not contaminated[a]

	True and super efficiencies	
	Mean	Median
Pearson correlation coefficient	−0.213	−0.196
Spearman correlation coefficient	−0.113	−0.117

[a] Results are for observations that have BCC efficiency = 1 and feasible solutions from super-efficiency model.

Table 2
Results of regressing pearson correlation coefficients between the true and super-efficiencies on explanatory variables when data are not contaminated[a]

Explanatory variables	Pearson correlation coefficient between true and super-efficiencies	
	Parameter estimates	t-Values
Intercept	−0.1949	−0.692
N	−0.0013	−3.462***
β	−0.0177	−0.067
$\bar{u}$	0.1015	0.985
R^2	0.013	

Pearson correlation coefficient $= \gamma_0 + \gamma_1 n + \gamma_2\beta + \gamma_3\bar{u} + \varepsilon$.
N is the sample size, $\beta = \beta_1 + \beta_2$ measures the concavity of the production function, and $\bar{u}$ is the mean inefficiency.
* Indicates significant at 10% level, ** indicates significant at 5% level, *** indicates significant at 1% level.
[a] Results are for observations that have BCC efficiency = 1 and feasible solutions from super-efficiency model.

[5] Each linear program in our simulation experiments took about 1 min to solve using the SAS software package under the UNIX operating system on Apache.

Table 3
Means and medians of correlation coefficients between the true and estimated efficiencies when data are not contaminated

	True and BCC efficiencies Mean (median)	True and super-efficiencies Mean (median)
Pearson correlation coefficient	0.917 (0.926)	0.889 (0.905)
Spearman correlation coefficient	0.883 (0.891)	0.894 (0.901)

Collectively, the simulation results reported in Tables 1–3 indicate that the super-efficiency scores are not appropriate for ranking efficient units as described in the AP procedure.

4. Outlier identification

The AP ranking procedure does not consider the potential impact of outliers on efficiency estimation. Outliers are a few extreme observations often caused by errors in measuring either the inputs or outputs. Since extreme observations determine the production frontier in DEA models, the estimation of the frontier may be sensitive to measurement errors in the sample data. If an observation has been contaminated with noise that increases the observed output value or decreases the observed input values such that it gets rated as efficient, then it may also enter the reference set of other observations and distort their estimated efficiency scores. Such outliers may be influential in the estimation results obtained using a conventional DEA model. It is desirable, therefore, to consider a procedure that allows us to identify and remove such outliers.

Banker and Gifford's (1988) procedure for identifying outliers generalizes Timmer's (1971) procedure. Timmer suggests discarding a certain percentage of efficient observations from the sample and re-estimating the production frontier using the remaining observations. Another way to interpret Timmer's procedure is that a certain proportion of efficient observations are classified as outliers and eliminated before re-estimating the efficiency of the remaining observations. BG's procedure differs from that of Timmer's in that they suggest the use of a screen based on the super-efficiency score to identify those observations that are more likely to be contaminated with noise. In other words, rather than throwing out an arbitrary set of efficient observations, BG suggest that only those observations with super-efficiency scores higher than a pre-selected screen should be eliminated. If an efficient observation is an outlier that has been contaminated with noise then it is more likely to have an output (or input) level much greater (smaller) than that of other observations with similar input (or output) levels. Therefore, such outliers are more likely to have a super-efficiency score much greater than one. This is the motivation underlying the BG procedure for outlier identification (see Banker et al., 1989, pp. 279–280).

4.1. Data generating process

To evaluate the BG procedure for outlier identification, we extended the data generating process described earlier for the simulation experiments in the previous section.

We considered two additional factors: probability that data are contaminated and the distribution of the random noise for such contaminated observations.

4.1.1. Probability of contaminated observations

The probability of an observation being contaminated with random noise was specified to be ρ. We generated ρ randomly from a uniform probability distribution over the interval $[0, 0.1]$. In other words, the probability that an observation is contaminated with random noise ranges from 0% to 10%.

4.1.2. Random noise distribution

Conditional on an observation being contaminated, we specified a two-sided random noise distribution. We generated the logarithm of the random noise v_k for each observation $k = 1, \ldots, N$, from a normal distribution $N(0, \sigma_v^2)$, where $\sigma_v = \delta E(z)$. For each experiment, the parameter δ was generated randomly from a uniform probability distribution over the interval $[0,1]$ and $E(z)=\frac{(15^{1+\beta_1}-5^{1+\beta_1})(15^{1+\beta_2}-5^{1+\beta_2})}{100(1+\beta_1)(1+\beta_2)}$ equaled the mean of the efficient output quantity.

Although we consider a two-sided random noise distribution, only those outliers that lie above the frontier are likely to affect the efficiency estimation of other observations by entering their reference set. Outliers with negative errors are likely to lie inside the frontier and have no impact on the efficiency estimation of other observations.

4.1.3. Simulated observations

For each experiment, we first generated values of N, β_1, β_2 and σ_u^2 as described in the previous section. In addition, we simulated values of ρ, the probability of being contaminated with random noise, from a uniform distribution over $[0, 0.1]$, and δ from a uniform distribution over $[0,1]$ to specify the parameter of the random noise distribution as $\sigma_v = \delta E(z)$. Next, for each observation k in the sample for an experiment, we randomly generated the values of the input quantities x_{1k} and x_{2k}, and inefficiency $\theta_k = \exp(u_k)$ as described in the previous section. Further, for each observation k, we also generated an index variable, q, from a uniform distribution over the interval $[0,1]$ and random noise v_k from the half-normal distribution $|N(0,\sigma_v^2)|$. Finally, we obtained the values for the observed output quantities y_k as

$$\begin{aligned}&\text{either } y_k = f(x_{1k},x_{2k}) * \exp(v_k)/\exp(u_k) \quad \text{if } q \leqslant \rho\\&\text{or } y_k = f(x_{1k},x_{2k})/\exp(u_k) \quad \text{if } q > \rho.\end{aligned} \tag{6}$$

4.2. Screens for outlier identification

To evaluate the performance of the BG procedure, we considered four different screen levels for outlier identification: 1.0, 1.2, 1.6 and 2.0. The screen level of 1.0 implies the elimination of *all* observations rated as efficient in the initial evaluation using the BCC model, while the other three screens require the solutions of the BG super-efficiency model. In the first stage, we identify and eliminate outliers using these pre-selected screen levels and then in the second stage re-estimate the BCC models with the remaining observations. We refer to the second-stage efficiency estimates as the BG-SE estimates.[6]

4.3. Simulation results: Average performance

In Panel A of Table 4 we report the mean and median Pearson and Spearman correlation coefficients between the true efficiency values and the BG efficiency estimates for each of the four different screen levels based on super-efficiency estimates. For comparison purposes, we also report the mean and median correlation coefficients between the true efficiency scores and the BCC efficiency estimates.

The results presented in Panel A of Table 4 indicate that the performance of the BG procedure using screens based on super-efficiency is better than that of the initial BCC estimates. With a screen of 1.2, the proportionate improvement in the Pearson correlation coefficient resulting from the use of the BG procedure is about 53% (=(0.490 − 0.320)/0.320) for the mean correlation and about 243% (=(0.559 − 0.163)/0.163) for the median correlation.[7] The performance of the BG procedure improves with an increase in the screen level to 1.2 and deteriorates when the screen level is increased further to 1.6. That is, the use of a more stringent screen level such as 1 is likely to misclassify many uncontaminated efficient observations as outliers, while the use of a less stringent screen level such as 1.6 or greater may fail to remove many contaminated observations.

The performance of the BG procedure analyzed by the ex ante probability of data contamination is reported in Panel B of Table 4. The performance of both the BCC model and the BG procedure deteriorates with the increase in the probability

[6] This second stage estimation procedure involves the solution of additional 8 (=4 + 4) linear programs, which required, on average, a total of about 8 min in our computational environment.

[7] The mean Pearson correlation coefficient between the true and BCC efficiency estimates is much greater than the median Pearson correlation coefficient due to a skewed distribution of correlation coefficient estimates: There are 629 out of the 1000 samples that have a Pearson correlation coefficient less than the overall mean Pearson correlation coefficient for all 1000 samples.

EOR 7001 ARTICLE IN PRESS No. of Pages 10, DTD = 5.0.1
4 August 2005 Disk Used

Table 4
Means and medians of correlation coefficients between the true efficiencies and BG estimated efficiencies based on super-efficiency model when data are contaminated

Procedure	BG-SE				BCC
Screen	1.0 Mean (median)	1.2 Mean	1.6 Mean (median)	2.0 Mean (median)	None Mean (median)
Panel A: Correlation coefficients for all simulations, using BG screens based on super-efficiency (BG-SE)					
Pearson correlation coefficient	0.482 (0.556)	0.490 (0.559)	0.482 (0.540)	0.469 (0.514)	0.320 (0.163)
Spearman Correlation coefficient	0.590 (0.690)	0.600 (0.705)	0.593 (0.690)	0.581 (0.652)	0.477 (0.411)

Procedure		BG-SE				BCC
Screen		1.0	1.2	1.6	2.0	None
Probability of contamination	Correlation coefficient					
Panel B: Mean correlation coefficients by screen level, using BG screens based on super-efficiency (BG-SE)						
0.00–0.01	Pearson	0.837	0.865	0.861	0.849	0.744
	Spearman	0.816	0.853	0.850	0.837	0.769
0.01–0.02	Pearson	0.778	0.799	0.798	0.786	0.560
	Spearman	0.809	0.835	0.829	0.812	0.661
0.02–0.03	Pearson	0.597	0.625	0.625	0.601	0.425
	Spearman	0.684	0.713	0.714	0.692	0.566
0.03–0.04	Pearson	0.525	0.533	0.519	0.515	0.329
	Spearman	0.626	0.634	0.620	0.617	0.473
0.04–0.05	Pearson	0.502	0.510	0.504	0.495	0.288
	Spearman	0.625	0.633	0.630	0.623	0.478
0.05–0.06	Pearson	0.417	0.429	0.420	0.408	0.241
	Spearman	0.574	0.584	0.575	0.563	0.438
0.06–0.07	Pearson	0.325	0.321	0.305	0.287	0.145
	Spearman	0.471	0.461	0.451	0.433	0.326
0.07–0.08	Pearson	0.284	0.283	0.277	0.252	0.161
	Spearman	0.432	0.435	0.426	0.409	0.341
0.08–0.09	Pearson	0.284	0.274	0.254	0.244	0.145
	Spearman	0.440	0.435	0.418	0.410	0.353
0.09–0.10	Pearson	0.262	0.252	0.246	0.240	0.150
	Spearman	0.420	0.414	0.409	0.405	0.354

of contamination. Similar to the overall results reported in Panel A of Table 4, the performance of the BG procedure improves with the increase in the screen level up to 1.2 when the probability of contamination is less than 0.06. For higher level of probability of contamination, the screen level 1.0 yields the best results as there are more contaminated observations. The most striking result, however, is that there is considerable improvement when using the BG procedure relative to the BCC model even for very small probabilities of contamination.

To further assess how well the BG procedure performs in efficiency estimation when the data are *not* contaminated with random noise, we report in Table 5 correlation coefficients between the true efficiency values and the BG estimates, and the true efficiency values and the BCC estimates. These results in Table 5 confirm the intuitive expectation that the BG estimation

Table 5
Means and medians of correlation coefficients between the true efficiencies and BG estimated efficiencies when data are not contaminated[a]

Procedure	BG-SE				BCC
Screen	1.0 Mean (median)	1.2 Mean (median)	1.6 Mean (median)	2.0 Mean (median)	None Mean (median)
Pearson correlation coefficient	0.864 (0.878)	0.898 (0.907)	0.907 (0.917)	0.908 (0.918)	0.917 (0.926)
Spearman correlation coefficient	0.812 (0.827)	0.859 (0.868)	0.869 (0.879)	0.870 (0.881)	0.883 (0.891)

[a] Correlation coefficients for all simulations, using BG screens based on super-efficiency (BG-SE).

procedure performs worse than the BCC procedure in this case, and that there is a gradual improvement in the performance when the screen level is made less stringent. This implies that the removal of some observations by using the BG procedure is likely to result in worse, rather than better, estimation performance if there is no noise contamination. Under such circumstances, there is a cost to eliminating observations identified as outliers using the BG procedure.

4.4. Simulation results: Regression analysis

To evaluate factors driving differences in performance across all experiments both when data are contaminated with random noise and when data are not contaminated with random noise, we specify the following regression model:

$$\Delta = \gamma_0 + \gamma_1 N + \gamma_2 \beta + \gamma_3 \bar{u} + \gamma_4 \text{ DUM1.2} + \gamma_5 \text{ DUM1.6} + \gamma_6 \text{ DUM2.0} + \varepsilon, \qquad (7)$$

where $\Delta = \text{Corr (True, BG)} - \text{Corr (True, BCC)}$, Corr (True, BG) is the Pearson correlation coefficient between the true efficiency and the BG estimates, Corr (True, BCC) is the Pearson correlation coefficient between the true efficiency and the BCC estimates, ρ is the probability of contamination of data with random noise, σ_v is the standard deviation of the normally distributed random noise, DUM1.2 is the dummy for screen level 1.2, DUM1.6 is the dummy for screen level 1.6 and DUM2.0 is the dummy for screen level 2.0. As before, N is the sample size, $\beta = \beta_1 + \beta_2$ is a measure of the concavity of the production function and $\bar{u} = E(e^u)$ is the mean inefficiency.

The regression results are reported in Table 6. As may be expected, the improvement in the correlation between the true and BG estimates based on super-efficiency increases significantly with the probability of data being contaminated with random noise, with the mean of the true inefficiency, and with the standard deviation of random noise.

Table 6
Regression results of differences in Pearson correlation coefficients between the True and BG, and True and BCC estimated efficiencies for all simulations both when data are contaminated and when data are not contaminated

Explanatory variables	Corr (True, BG-SE)–Corr (True, BCC)	
	Parameter estimates	t-Values
Intercept	−0.1289	−2.14**
N	−0.0002	−2.11***
β	0.0326	0.58
ρ	0.7495	8.50***
$\bar{u}$	0.0817	3.73***
σ_v	0.0245	22.03***
DUM1.2	0.0212	3.29***
DUM1.6	0.0213	3.30***
DUM2.0	0.0150	2.33**
R^2	0.135	

$\Delta = \gamma_0 + \gamma_1 N + \gamma_2 \beta + \gamma_3 \rho + \gamma_4 \bar{u} + \gamma_5 \sigma_v + \gamma_6 \text{DUM1.2} + \gamma_7 \text{DUM1.6} + \gamma_8 \text{DUM2.0} + \varepsilon$.
$\Delta = \text{Corr (True, BG)} - \text{Corr (True, BCC)}$, N is the sample size, $\beta = \beta_1 + \beta_2$ measures the concavity of the production function, ρ is the probability of data contamination, $\bar{u}$ is the mean inefficiency, σ_v is the standard deviation of the random noise, DUM1.2 is the dummy for screen level 1.2, DUM1.6 is the dummy for screen level 1.6 and DUM2.0 is the dummy for screen level 2.0.
* Indicates significant at 10% level.
** Indicates significant at 5% level.
*** Indicates significant at 1% level.

While the degree of production function concavity β does not have a significant impact on performance measured in terms of the difference in the correlation coefficients between the true and BG estimates and between the true and BCC efficiencies, the difference in the two correlation coefficients decreases significantly with the sample size. This is because the BCC model is more likely to retrieve the true efficiency when the sample size becomes larger due to the consistency property of the DEA estimator (Banker, 1993), while the performance of the super-efficiency measure deteriorates with sample size as observed earlier in the regression results reported in Table 2. Finally, the results in Table 6 also indicate that the coefficient estimates of all three dummies for screen levels 1.2, 1.6 and 2.0 are significantly positive for the super-efficiency based screens.

5. Conclusion

In this paper, we have conducted simulation experiments to evaluate the performance of the Banker and Gifford (1988) super-efficiency model when it is used for ranking efficient units and when it is used for outlier identification. We find that Andersen and Petersen's (1993) procedure using the super-efficiency model for ranking efficient observations does not perform satisfactorily. In fact, correlation coefficients between the true efficiency values and the estimated super-efficiency scores of efficient observations are negative in a majority of the cases. In contrast, the evidence seems to support the use of Banker and Gifford's (1988) and Banker et al.'s (1989) super-efficiency based procedure to identify outliers. This method provides more accurate efficiency estimates than the BCC estimates when data are contaminated with random noise.

References

Andersen, P., Petersen, N., 1993. A procedure for ranking efficient units in data envelopment analysis. Management Science 39, 1261–1264.

Banker, R.D., 1993. Maximum likelihood, consistency and data envelopment analysis: A statistical foundation. Management Science (October), 1265–1273.

Banker, R.D., Charnes, A., Cooper, W.W., 1984. Models for the estimation of technical and scale inefficiencies in data envelopment analysis. Management Science (September), 1078–1092.

Banker, R.D., Gifford, J.L., 1988. A relative efficiency model for the evaluation of public health nurse productivity, Mellon University Mimeo, Carnegie.

Banker, R.D., Das, S., Datar, S.M., 1989. Analysis of cost variances for management control in hospitals. Research in Governmental and Nonprofit Accounting 5, 269–291.

Charnes, A., Cooper, W.W., Rhodes, E., 1978. Measuring the efficiency of decision making units. European Journal of Operational Research 2, 429–444.

Charnes, A., Neralic, L., 1990. Sensitivity analysis of the additive model in data envelopment analysis. European Journal of Operational Research 48, 332–341.

Charnes, A., Neralic, L., 1992. Sensitivity analysis of the proportionate change of inputs (or outputs) in data envelopment analysis. Glasnik Matematicki 27, 393–405.

Chen, Y., 2000. On ranking efficient units. Merrimack College, Mimeo.

Seiford, L.M., Zhu, J., 1999. Infeasibility of super-efficiency data envelopment analysis. INFOR 37, 174–187.

Timmer, C.P., 1971. Using a probabilistic frontier production function to measure technical efficiency. Journal of Political Economy 79, 776–794.

Wilson, P., 1995. Detecting influential observations in data envelopment analysis. Journal of Productivity Analysis 6, 27–45.