CrossMark

On the use of super-efficiency procedures for ranking efficient units and identifying outliers

Rajiv D. Banker[1] · **Hsihui Chang**[2] ·
Zhiqiang Zheng[3]

Abstract Prior research (Banker and Chang in Eur J Oper Res 175: 1311–1320, 2006) has found that super-efficiency based procedures are effective in identifying outliers, but not in ranking efficient units. In this paper, we investigate why the procedures failed to rank efficient units satisfactorily and examine the performance of super-efficiency procedures in different "regions" of production set. We find that the unsatisfactory results mainly originate from the "left corners" of DMUs, those units with relatively smaller values. We further examine the effect of different noise levels on outlier detection using the BG procedure (Banker and Gifford in A relative efficiency model for the evaluation of public health nurse productivity. Mimeo, Carnegie Mellon University, Pittsburgh, 1988). Our results show that the BG procedure is more effective when the noise level is high. We conduct extensive simulation experiments by considering different production functions (Cobb–Douglas and Polynomial), different DEA formulations (BCC and CCR) and different returns-to-scale (CRS and NIRS) assumptions. Our simulation results confirm that the findings in Banker and Chang (2006) are robust under different DEA formulations, production functions and returns to scale assumptions.

Keywords DEA · Super-efficiency · Outlier identification · Efficiency ranking · Simulation study

✉ Hsihui Chang
  hc336@drexel.edu

[1] Fox School of Business, Temple University, Philadelphia, PA, USA

[2] LeBow College of Business, Drexel University, Philadelphia, PA, USA

[3] Jindal School of Management, University of Texas at Dallas, Richardson, TX, USA

1 Introduction

In the DEA literature, the super-efficiency procedure has been proposed to rank efficient decision making units (DMUs) by Andersen and Petersen (1993, AP thereafter) and to identify outliers (Banker and Gifford 1988; Banker et al. 1989). Under the super-efficiency DEA formulation, the efficiency scores may be different and do not have to equal one for those efficient DMUs. Thus the AP procedure can be suggested to discriminate those efficient units. Nevertheless there lacks evidence on whether the super-efficiency model is effective to differentiate those efficient units despite its popularity in DEA applications. Banker and Chang (2006) find that the AP procedure does not rank efficient units satisfactorily in their simulation study but they do not elaborate on why the AP procedure is ineffective in ranking efficient units. This paper attempts to shed light on this problem by examining the performance of super-efficiency procedures in different "regions" of the production set, i.e. for DMUs with different levels of input/output values. We find that the unsatisfactory results mainly originate from the "left corners" of DMUs, those units with relatively smaller values.

In the case of outlier detection, Banker and Gifford (1988) suggest using super-efficiency scores to screen out those observations whose super-efficiency scores exceed a pre-specified screen level (see Banker et al. 1989, pp. 279–280).[1] After removing those outliers, a conventional DEA model (e.g. the BCC model) is estimated with the remaining observations. We refer to this approach as the BG procedure hereafter. Banker and Chang (2006) show that the BG procedure is effective in identifying outliers in the presence of noise data. This finding is consistent with our results based on extensive simulation experiments. We further investigate the effect of different noise levels on outlier detection using the BG procedure. We find that the BG procedure is generally more effective when the noise level is high.

It is important to note that the simulation study in Banker and Chang (2006) only considers a Cobb–Douglas production function and a super-efficiency procedure for the BCC model. It remains to be seen if their findings are consistent across different production functions, different DEA formulations and different returns-to–scale settings. Our experiments extend Banker and Chang (2006) in three dimensions: (1) we incorporate different production functions based on a polynomial function that displays different characteristics (Banker et al. 2004, Banker and Chang 1996); (2) we examine the performance of super-efficiency procedures under both BCC and CCR formulations; and (3) we consider different returns-to-scale (CRS and NIRS) assumptions. Our simulation results show that the findings in Banker and Chang (2006) are robust to different DEA formulations, production functions and returns-to-sscale assumptions.

The remainder of this paper is organized as follows. We first describe the super-efficiency model in DEA in Sect. 2. Section 3 details the data generating process and the simulation setup for testing the performance of the AP procedure on ranking efficient units. Section 4 describes the data generating process and simulation results to evaluate the performance of the super-efficiency model in outlier identification. Section 5 concludes with a summary of our principal results.

2 Super-efficiency models

We consider three specific super-efficiency models: super BCC, super CCR, and super NIRS models.

---

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

2.1 Super BCC

We refer to the approach of using the super-efficiency procedure on the BCC formulation as the super BCC model. Let $Y_j = 0$ and $X_j = 0$, $j = 1, \ldots N$, 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 $\hat{\psi}_k^{SE}$ for an observation (X_k, Y_k), $k \in \{1 \ldots N\}$ is the reciprocal of the super-inefficiency measure $\hat{\theta}_k^{SI}$ obtained by solving the following linear program:

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

subject to

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

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

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

$$\theta_k, \lambda_j \geq 0 \tag{1d}$$

Note that the above super-efficiency model excludes the observation "k" under evaluation in the reference set for the constraints in (1a), (1b) and (1c), as opposed to the conventional BCC model. Thus the reference observation $(\sum_{\substack{j=1 \\ j \neq k}}^{N} \lambda_j X_j, \sum_{\substack{j=1 \\ j \neq k}}^{N} \lambda_j Y_j)$ in the evaluation of the super-efficiency of observation k is constructed only from observations other than k itself. Nevertheless this does not guarantee a convex combination can be created from the remaining observations to envelop observation k. 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:

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

---

[2] While the super-efficiency model is specified in (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 NDRS (non-decreasing 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 (=) constraint instead of an equality (=) constraint. The NIRS (non-increasing 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 (=) constraint instead of an equality (=) constraint.

subject to

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

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

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

$$\eta_k, \lambda_j \geq 0 \tag{2d}$$

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 $\hat{\psi}_k^{SE} = 1/\eta_k^*$ 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 $\hat{\psi}_k^{SE} < 1$, and that the observation would have been rated as efficient by the conventional BCC model if and only if $\hat{\psi}_k^{SE} \geq 1$ (Banker and Gifford 1988).

2.2 Super CCR and super NIRS

Similar to the above Super BCC model, the Super CCR model excludes a unit "k" from the reference set in evaluating the super-efficiency of a certain observation. However in the Super CCR model, Eq. (2c) is omitted because of the constant returns to scale assumption in the CCR formulation.

The Super NIRS model assumes a non-increasing returns to scale, that is, $\sum_{j=1}^{N} \lambda_j \geq 1$. The only difference between the Super NIRS formulation and the Super BCC formulation is that the former replaces Eq. (2c) in the latter with $\sum_{j=1}^{N} \lambda_j \geq 1$.

3 Ranking efficient units in different "regions"

As mentioned earlier, Andersen and Petersen (AP) suggested the use of the super-efficiency model for ranking efficient units. This section aims to evaluate the performance of the AP procedure in ranking efficient units and more importantly to shed light on when the AP procedure is effective or ineffective. We conduct 500 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] Observe that there always exists an optimal solution to (2) with either $\lambda_k^* = 0$ or $\lambda_k^* = 1$.

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 two different production technologies. The first consists of a single output and two inputs specified in terms of its efficient 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 generated randomly from independent uniform probability distributions over the interval [10, 20], and the coefficients β_1 and β_2 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. The function shows 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}, $i = 1, 2$.

   The second production technology we consider here follows Banker and Natarajan (2004). It is a polynomial function in the form of

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 \tag{4}$$

The input variable x is generated from the uniform distribution over the interval [1,4]. The coefficients $\beta_0, \beta_1, \beta_2, \beta_3$ determine the properties of the production of technology and we use the following values: $\beta_0 = -37, \beta_1 = 48, \beta_2 = -12, \beta_3 = 1$. Note that these choices guarantee that the production function is a monotonically increasing and concave function when $x \in [1, 4]$.

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 $\left|N(0, \sigma_u^2)\right|$, 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 Cobb–Douglas production function

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

---

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

(x_{1k}, x_{2k}), k = 1, …N, were then substituted into the efficient production function specified in Eq. (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 ∈ {1, …N} from the half-normal distribution $\left|N(0, \sigma_u^2)\right|$. Finally, we obtained the values for "observed" output quantities y_k and the "true" efficiency values ψ_k as:

$$y_k = f(x_k)/\exp(u_k) \tag{5}$$

$$\psi_k = 1/\exp(u_k). \tag{6}$$

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.

3.1.5 Simulated observations for the polynomial production Function

Similarly, for each experiment we first randomly generated a value for N between 51 and 150 and a value for σ_u^2 between 0 and 0.1989. Next, we simulated N observations of the input x between a and 4. These values x_k, k = 1, …N, were then substituted into the efficient production function specified in Eq. (4) to obtain the corresponding values $z_k = f(x_k)$ for the efficient output quantity. Then, we randomly generated the logarithm of "true" inefficiency values $u_k = \theta_k$ for each observation k ∈ {1, …N} from the half-normal distribution $\left|N(0, \sigma_u^2)\right|$. Finally, we obtained the values for "observed" output quantities y_k:

$$y_k = f(x_{1k}, x_{2k}) - u_k \tag{7}$$

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

3.2 Regions

We hypothesize that efficiency rankings derived from super-efficiency procedures are more sensitive to those units on the corners than those in the middle. To test this hypothesis, we separate the DMUs into three regions based on the values of inputs/outputs: 15 % to the left as region 1, 70 % in the middle as region 2 and 15 % to the right as region 3.

Superficially, we divide the DMUs generated from the Cobb–Douglas production function (see Eq. 5) based on the value of y_k. Those 15 % DMUs with the smallest y_k are assigned to region 1; the 15 % with the highest y_k are assigned to region 3; and the rest middle 70 % are assigned to region 2. This roughly corresponds to the regions having $y_k = 5$, $5 < y_k = 8$ and $8 < y_k$ for region 1, 2 and 3 respectively.

For the polynomial function, we separate the three regions according to y_k specified in Eq. (7) in a similar manner—15 % DMUs with the smallest y_k are assigned to region 1, 15 % with the highest y_k are assigned to region 3, and the rest middle 70 % are assigned to region 2. This roughly corresponds to the three regions of $y_k = 11.2$, $11.2 < y_k = 25.2$ and $25.2 < y_k$ respectively.

3.3 Simulation results

In each experiment, we run six linear programs in order to estimate the efficiencies for the six models considered in this paper: BCC, CCR, NIRS, the Super-BCC, Super CCR and Super-NIRS, for each of the k=1,…N observations in the sample.

In Table 1, we report the means of Pearson correlation between the true efficiency value ψ_k and the estimated efficiency $\hat{\psi}_k$. It is evident from Panel A and B of Table 1 that AP

Table 1 Means of pearson correlation coefficients between the true and estimated efficiencies

	Super-BCC	Super-CCR	Super-NIRS			
Panel A: correlation based on all observations that have BCC/CCR/NIRS efficiency = 1						
Cobb–Douglas	−0.434	0.134	−0.454			
Polynomial	−0.222	0.180	−0.202			
	BCC	CCR	NIRS	Super-BCC	Super-CCR	Super-NIRS
Panel B: correlation based on all observations						
Cobb–Douglas	0.935	0.912	0.934	0.637	0.911	0.626
Polynomial	0.659	0.177	0.662	0.388	0.177	0.391

procedure's performance in ranking efficient units is not at all satisfactory, consistent with the findings of Banker and Chang (2006). One claimed advantage of the AP procedure is that it can distinguish efficiencies among those efficient observations. Our results in Panel A did not lend support to this. Panel A shows that the correlations (only for those efficient points) are very low and many of them are even negative, across all six models considered. Panel B shows that the mean correlations (based on all observations) of AP's procedures (super-BCC, super-CCR and super-NIRS) are all lower than those of their counterparts (BCC, CCR and NIRS) across both production functions (Cobb–Douglas and Polynomial). In addition, while CCR and super CCR seem to be appropriate for the Cobb–Douglas production function, they are not satisfactory for the polynomial function. This may be because the concavity of the polynomial function violates the constant returns to scale assumption of CCR.

What might have caused the above findings? We hypothesize that the performance of AP's procedure may vary in different regions of observations. Table 2 summarizes the comparative results across the three regions of DMUs.[5] Panel A shows that the AP procedure does not rank those efficient units well for any of the three regions, with region 1 being the worst, followed by region 2 and then region 3. The results in Panel B and C of Table 2 are based on all observations. These two panels show that both the AP procedure (see Panel B) and the conventional DEA models (see Panel C) perform the worst in region 1. However, the problem is more pronounced for the AP procedure. In general, region 2 and region 3 results are more satisfactory for both the AP procedure and the conventional DEA models. Between the two panels, Panel C results are consistently better than those of Panel B across regions. This furthers corroborates that the AP procedure does not improve the rankings of DMUs over the conventional DEA formulations (BCC, CCR and NIRS).

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

---

[5] Table 2 has the same structure as Table 1, but with the results from each region. When interpreting Table 2 results, one should not expect the results of Table 1 as the average of the three corresponding regions in Table 2 because of different sample sizes.

Table 2 Means of pearson correlation coefficients between the true and estimated efficiencies when data are separated into regions

	Super BCC			Super CCR			Super NIRS		
	Region1	Region2	Region3	Region1	Region2	Region3	Region1	Region2	Region3
Panel A: super-efficiency model results based on observations that have BCC/CCR/NIRS efficiency = 1 within each region									
Cobb–Douglas	−0.553	−0.443	0.104	0.098	0.147	0.349	−0.591	−0.423	0.264
Polynomial	−0.393	0.033	0.224	0.002	0.180	0.875	−0.389	0.013	0.447
	Super BCC			Super CCR			Super NIRS		
	Region1	Region2	Region3	Region1	Region2	Region3	Region1	Region2	Region3
Panel B: super-efficiency model results based on all observations within each region									
Cobb–Douglas	0.634	0.757	0.879	0.883	0.895	0.918	0.647	0.776	0.950
Polynomial	0.527	0.919	0.973	0.240	0.335	0.493	0.530	0.918	0.973
	BCC			CCR			NIRS		
	Region1	Region2	Region3	Region1	Region2	Region3	Region1	Region2	Region3
Panel C: BCC, CCR and NIRS results based on all observations within each region									
Cobb–Douglas	0.858	0.959	0.904	0.883	0.897	0.936	0.838	0.961	0.961
Polynomial	0.864	0.912	0.973	0.226	0.327	0.473	0.868	0.912	0.972

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 (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

Noise distribution for the Cobb–Douglas function 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,…N, from a normal distribution $N(0, s_v^2)$, where $s_v = d\,E(z)$. For each experiment, the parameter d 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.

Noise distribution for the polynomial function We generated the random noise v_k for each observation k = 1, …N, from a half-Normal distribution $|N(0, \sigma_v^2)|$, where $\sigma_v = d/2\,E(z)$, where parameter d was generated randomly from a uniform probability distribution over the interval [0, 1] for each experiment and E(z) = 20.2.

4.1.3 Simulated observations

Simulated observations for the Cobb–Douglas function 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 d 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 $\left|N(0, \sigma_v^2)\right|$. Finally, we obtained the values for the observed output quantities y_k as:

$$
\begin{aligned}
&\text{either } y_k = f((x_{1k},\ x_{2k}) * \exp(v_k)/\exp(u_k) && \text{if } q \leq \rho\\
&\text{or } \quad\ \ y_k = f((x_{1k},\ x_{2k})/\exp(u_k) && \text{if } q > \rho\,.
\end{aligned}
\tag{8}
$$

Simulated observations for the polynomial function For each experiment, we first generated values of N, ρ and σ_u^2 as described in the previous section. Next, for each observation k in the sample for an experiment, we randomly generated the values of the input quantities x_k and inefficiency $\theta_k = u_k$ 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 $\left| N(0, \sigma_v^2) \right|$. Since DEA does not allow non-positive values, we added a constant, 10, to y_k to ensure all values are positive. Finally, we obtained the values for the observed output quantities y_k as:

$$\begin{array}{ll} \text{either} \quad y_k = f(x_k) - u_k - v_k + 10 & \text{if } q \leq \rho \\ \text{or} \quad\quad y_k = f(x_k) - u_k & \text{if } q > \rho . \end{array} \tag{9}$$

4.2 Screens for outlier identification

To evaluate the performance of the BG procedure, we considered the screen level 1.0 for outlier identification as suggested by Banker and Gifford (1988). The screen level of 1.0 implies the elimination of *all* observations rated as super-efficient in the BG super-efficiency model. In the first stage, we identify and eliminate outliers using the pre-selected screen level and then in the second stage re-estimate the BCC, CCR and NIRS models with the remaining observations. We refer to the second-stage efficiency estimates as the BG-SE estimates.

4.3 Simulation results: average performance

Table 3 reports the mean correlations between the true efficiency values and the BG efficiency estimates. For comparison purposes, we also report the mean correlations between the true efficiency scores and the efficiency estimates from the BCC, CCR and NIRS models. The main finding here is that the BG procedure is effective in identifying outliers, across both production functions and the three different DEA models. On average, the results in Table 3 indicate that the BG procedures using screens outperforms all of the initial BCC, CCR and NIRS estimators. For example with a screen of 1.0, using the BG super efficiency procedure improves ovesr the BCC model by 19 % (= (0.635−0.543)/0.543), for the Cobb–Douglas function averaged across different noise levels. Interestingly there is considerable improvement when using the BG procedures relative to the conventional DEA models even for very small probabilities of contamination. Figure 1 is based on the BG procedure results in Table 2, with Cobb–Douglas results on the left chart and the polynomial results on the right one. Clearly, as the noise level increases, the performance of the BG procedures deteriorates, for all three DEA models we consider. In addition, the BG procedure for the CCR model does not perform well for the polynomial function. As mentioned earlier, this may be due to the fact that the constant returns to scale assumption of CCR is violated under the polynomial function.

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 4 correlation between the true efficiency values and the BG estimates, as well as the correlation between true efficiency values and the BCC/CCR/NIRS estimates. In general, these results in Table 4 confirm the intuition that the BG estimation procedure may not be as effective in this case. Out of six scenarios in Table 4, only two (Super-BCC and Super-NIRS for the polynomial function) outperform their counterparts. The remaining four perform worse. This implies that the removal of some observations by using the BG procedure is likely to result in worse,

Table 3 Correlation coefficients between the true efficiencies and BG estimated efficiencies based on super-efficiency model when data are contaminated, by Rho and screen levels

|	Screen level	Rho	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	Average
Cobb–Douglas	None	BCC	0.87	0.80	0.70	0.54	0.54	0.46	0.43	0.40	0.42	0.30	0.543
	None	CCR	0.84	0.84	0.81	0.68	0.69	0.65	0.61	0.56	0.55	0.47	0.669
	None	NIRS	0.90	0.77	0.68	0.64	0.64	0.48	0.44	0.38	0.40	0.40	0.573
	1.0	Super-BCC	0.83	0.83	0.82	0.66	0.66	0.61	0.54	0.52	0.50	0.38	0.635
	1.0	Super-CCR	0.82	0.84	0.82	0.71	0.70	0.66	0.61	0.57	0.57	0.46	0.676
	1.0	Super-NIRS	0.87	0.89	0.76	0.74	0.73	0.60	0.56	0.45	0.46	0.50	0.657
	None	BCC	0.72	0.71	0.71	0.69	0.28	0.26	0.26	0.28	0.27	0.28	0.445
	None	CCR	0.18	0.18	0.19	0.18	0.18	0.19	0.19	0.20	0.20	0.20	0.189
Poly-nomial	None	NIRS	0.71	0.71	0.70	0.54	0.26	0.26	0.27	0.26	0.27	0.27	0.425
	1.0	Super-BCC	0.77	0.75	0.74	0.71	0.27	0.26	0.26	0.27	0.27	0.28	0.457
	1.0	Super-CCR	0.05	0.05	0.06	0.06	0.16	0.15	0.16	0.17	0.16	0.16	0.117
	1.0	Super-NIRS	0.77	0.75	0.75	0.53	0.26	0.26	0.26	0.26	0.26	0.27	0.436
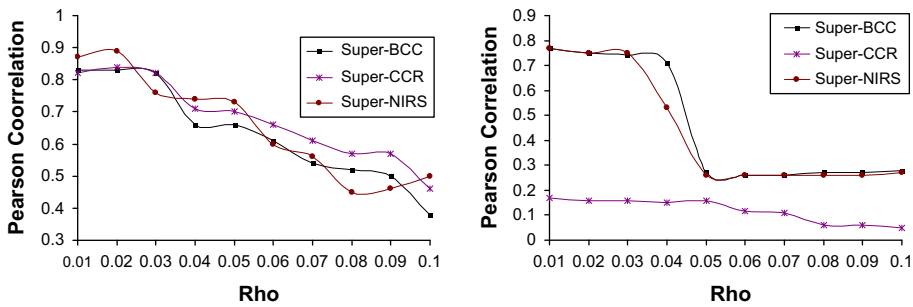
Fig. 1 Pearson correlation comparison as a function of noise rate rho under Cobb–Douglas (*left*) and Polynomial (*right*) production functions

Table 4 Means of correlation coefficients between the true efficiencies and BG estimated efficiencies when data are not contaminated

	Procedure	Super-efficiency	BCC	CCR	NIRS
Cobb–Douglas	Screen level	1.0	None	None	None
	Super-BCC	0.842	0.894	–	–
	Super-CCR	0.849	–	0.864	–
	Super-NIRS	0.862	–	–	0.957
Polynomial	Super-BCC	0.767	0.725	–	–
	Super-CCR	0.063	–	0.163	–
	Super-NIRS	0.765	–	–	0.692

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.

In order to shed light on the performance of BG procedures in different regions, Table 5 details correlations of different DEA procedures by region, noise level and the screen level. There are three main findings. First, all procedures' performance deteriorates for all regions as the noise level increases. Second, for the Cobb–Douglas function, the BG procedure shows clear improvement over the counterpart DEA model for each of the three regions. For example, for region 2, the BG procedure on top of the BCC model (referred to as super-BCC in Table 5) outperforms the BCC model by 21 % = (0.84-0.69)/0.69. However, for the polynomial functions, the BG-procedure results do not improve over its counterparts as significantly when the DMUs are separated into three regions. Third, most interestingly, as the noise level increases for the Cobb–Douglas function, only the efficiencies of those DMUs in the middle (region 2) are estimated well. Both of the two corner DMU regions (region 1 and region 3) are not estimated satisfactorily. This finding is consistent across all the six procedures (BCC, CCR, NIRS, Super-BCC, Super-CCR and Super-NIRS). We believe this is because the positive noise data in the middle region tend to be muted by the greater likelihood of finding spanning observations in constructing a convex combination of reference points, but this is less likely in the corners.

Table 5 Correlation coefficients between the true efficiencies and BG estimated efficiencies based on super-efficiency model when data are contaminated, by Rho, Screen levels and regions

Rho	Procedure	Region	Screen	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10	Average
Cobb–Douglas	BCC	1	None	0.69	0.70	0.53	0.43	0.41	0.20	0.16	0.29	0.37	0.04	0.38
		2	None	0.85	0.84	0.82	0.73	0.63	0.67	0.60	0.60	0.58	0.60	0.69
		3	None	0.73	0.48	0.40	0.40	0.17	0.02	0.07	-0.10	-0.10	-0.20	0.19
	CCR	1	None	0.75	0.76	0.66	0.48	0.37	0.29	0.20	0.30	0.25	0.14	0.42
		2	None	0.85	0.85	0.83	0.81	0.72	0.81	0.73	0.74	0.71	0.76	0.78
		3	None	0.76	0.49	0.47	0.39	0.12	-0.00	0.11	-0.10	-0.10	-0.20	0.19
	NIRS	1	None	0.79	0.67	0.46	0.42	0.54	0.35	0.30	0.36	0.18	0.15	0.42
		2	None	0.95	0.89	0.74	0.73	0.73	0.72	0.70	0.63	0.52	0.57	0.72
		3	None	0.93	0.71	0.20	0.31	0.28	0.27	0.01	-0.20	-0.20	-0.20	0.21
	Super-BCC	1	1.0	0.63	0.65	0.49	0.45	0.49	0.43	0.44	0.22	0.05	0.12	0.40
		2	1.0	0.92	0.93	0.86	0.90	0.86	0.85	0.85	0.76	0.75	0.71	0.84
		3	1.0	0.80	0.76	0.75	0.76	0.60	0.63	0.52	0.38	0.26	0.21	0.57
	Super-CCR	1	1.0	0.76	0.72	0.49	0.55	0.64	0.45	0.45	0.28	0.20	0.11	0.46
		2	1.0	0.89	0.87	0.82	0.85	0.84	0.83	0.85	0.77	0.77	0.78	0.83
		3	1.0	0.90	0.79	0.73	0.68	0.63	0.56	0.29	0.29	0.17	0.15	0.52
	Super-NIRS	1	1.0	0.70	0.68	0.58	0.55	0.53	0.43	0.31	0.27	0.13	0.16	0.43
		2	1.0	0.91	0.91	0.92	0.86	0.88	0.79	0.82	0.79	0.72	0.72	0.83
		3	1.0	0.86	0.94	0.92	0.66	0.78	0.55	0.51	0.54	0.35	0.29	0.64

Table 5 continued

Rho	Procedure	Region	Screen	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10	Average
Poly-nomial	BCC	1	None	0.92	0.92	0.92	0.92	0.53	0.46	0.43	0.47	0.43	0.46	0.65
		2	None	0.95	0.95	0.95	0.95	0.62	0.37	0.48	0.41	0.4	0.38	0.65
		3	None	0.99	0.87	0.78	0.79	0.77	0.75	0.73	0.75	0.76	0.69	0.79
	CCR	1	None	0.15	0.13	0.12	0.15	0.41	0.4	0.31	0.38	0.4	0.48	0.29
		2	None	0.21	0.22	0.21	0.21	0.20	0.15	0.18	0.17	0.19	0.2	0.19
		3	None	0.99	0.87	0.78	0.79	0.77	0.75	0.73	0.75	0.76	0.69	0.79
	NIRS	1	None	0.90	0.93	0.92	0.90	0.50	0.48	0.46	0.50	0.49	0.46	0.65
		2	None	0.94	0.96	0.95	0.95	0.65	0.37	0.40	0.38	0.41	0.41	0.64
		3	None	0.99	0.85	0.79	0.77	0.79	0.78	0.75	0.75	0.77	0.69	0.79
	Super-BCC	1	1.0	0.83	0.92	0.87	0.78	0.52	0.38	0.43	0.38	0.37	0.38	0.59
		2	1.0	0.93	0.94	0.94	0.91	0.79	0.28	0.19	0.25	0.33	0.30	0.59
		3	1.0	0.99	0.90	0.70	0.74	0.71	0.69	0.74	0.69	0.66	0.61	0.74
	Super-CCR	1	1.0	0.34	0.34	0.38	0.43	0.37	0.25	0.14	0.09	0.09	0.12	0.25
		2	1.0	0.12	0.12	0.12	0.10	0.10	0.08	0.04	0.05	0.11	0.09	0.09
		3	1.0	0.99	0.90	0.70	0.74	0.71	0.69	0.74	0.69	0.66	0.61	0.74
	Super-NIRS	1	1.0	0.83	0.86	0.87	0.61	0.41	0.44	0.41	0.42	0.43	0.41	0.57
		2	1.0	0.91	0.93	0.93	0.81	0.22	0.24	0.22	0.27	0.23	0.30	0.51
		3	1.0	0.99	0.80	0.73	0.73	0.69	0.71	0.71	0.69	0.75	0.60	0.74

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 contrast, the evidence supports the use of Banker and Gifford's (1988) and Bankera et al.'s (1989) super-efficiency based procedure to identify outliers. Most importantly, we document that the poor performance is particularly acute for observations with smaller input and output values. Performance is worse for the super-CCR model. Performance of the BG outlier screening procedure also deteriorates with the level of noise, and rather precipitously when the noise level rho exceeds 0.05.

Our study has the following important implications. From academic perspective, there is an urgent need for future research to propose alternative methods or to extend the AP procedure for the improvement of ranking the performance of efficient units. From practical consideration, given that the BG procedure outperforms other conventional DEA procedures only when the data is contaminated with a high level of random noise, it is essential for decision makers to evaluate the extent of their data contamination before using the BG procedure in removing outliers. However, a caveat is in order. The results obtained from this study may not be applied directly to the multiple output and multiple input case without conducting additional Monte Carlo simulation experiments.

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*, *39*, 1265–1273.

Banker, R. D., & Chang, H. (2006). The super-efficiency procedure for outlier identification, not for ranking efficient units. *European Journal of Operations Research*, *175*, 1311–1320.

Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Models for the estimation of technical and scale inefficiencies in data envelopment analysis. *Management Science*, *30*, 1078–1092.

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.

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

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.

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.