European Journal of Operational Research 240 (2015) 518–527

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Decision Support

The impact of mismeasurement in performance benchmarking: A Monte Carlo comparison of SFA and DEA with different multi-period budgeting strategies

Seog-Chan Oh [a,*], Jaemin Shin [b]

[a] General Motors Research and Development, 30500 Mound Road, Warren, MI 48090, United States
[b] Department of Mathematics, Inha University, Incheon 402-751, Republic of Korea

ARTICLE INFO

Article history:
Received 31 January 2013
Accepted 18 July 2014
Available online 1 August 2014

Keywords:
Efficiency estimation
Stochastic frontier analysis
Data envelopment analysis
Performance-based budgeting

ABSTRACT

Performance-based budgeting has received increasing attention from public and for-profit organizations in an effort to achieve a fair and balanced allocation of funds among their individual producers or operating units for overall system optimization. Although existing frontier estimation models can be used to measure and rank the performance of each producer, few studies have addressed how the mismeasurement by frontier estimation models affects the budget allocation and system performance. There is therefore a need for analysis of the accuracy of performance assessments in performance-based budgeting. This paper reports the results of a Monte Carlo analysis in which measurement errors are introduced and the system throughput in various experimental scenarios is compared. Each scenario assumes a different multi-period budgeting strategy and production frontier estimation model; the frontier estimation models considered are stochastic frontier analysis (SFA) and data envelopment analysis (DEA). The main results are as follows: (1) the selection of a proper budgeting strategy and benchmark model can lead to substantial improvement in the system throughput; (2) a "peanut butter" strategy outperforms a discriminative strategy in the presence of relatively high measurement errors, but a discriminative strategy is preferred for small measurement errors; (3) frontier estimation models outperform models with randomly-generated ranks even in cases with relatively high measurement errors; (4) SFA outperforms DEA for small measurement errors, but DEA becomes increasingly favorable relative to SFA as the measurement errors increase.

1. Introduction

Performance assessment of production and services is becoming an increasingly important managerial activity as public and for-profit organizations turn to performance-based budgeting in an effort to optimize the allocation of funds across their individual producers. The performance-based budgeting technique is clearly applicable and adaptable to many practical situations. One example is the allocation of an energy conservation program budget and assignment of appropriate energy quotas to the individual plants in a global automotive manufacturing company. The common practice in setting energy quotas is to apply the peanut butter approach,[1] in which the limited total budget is simply divided among individual plants in proportional to their production sizes along with an equal percentage of energy reduction quota (e.g., a 4% cut in the energy cost per production unit from the previous year), without considering the different energy saving potentials of the various plants. For example, certain plants may already have implemented aggressive energy saving programs and reached a point of diminishing returns, while other plants barely met (or even failed to meet) their allocated reduction quotas in previous years, carrying the remaining energy saving potential over to the next year. In such cases, the peanut butter approach may not serve to optimize the overall system performance or lead to efficient use of a limited budget. It would be far more desirable to allocate relatively low energy saving targets and low budgets to best-practice plants while allocating aggressive energy saving targets to inefficient plants along with high budgets to encourage major changes in their energy conservation practices.

The effectiveness of performance-based budgeting depends on the accuracy of the performance assessment in distinguishing between best-practice and inefficient producers. There are two major approaches to frontier estimation – stochastic frontier

* Corresponding author.
E-mail addresses: seog-chan.oh@gm.com (S.-C. Oh), jaemin.shin@inha.ac.kr (J. Shin).

[1] It is frequently used as a business term where it can refer to the efforts to apply the same tactics to all aspects or parts of a business. As a person makes a peanut butter sandwich by spreading the butter thin evenly on the bread, a business or a government may want to spread something (tactics, money, or tax break) evenly across all areas.

http://dx.doi.org/10.1016/j.ejor.2014.07.026
0377-2217/

analysis (SFA) and data envelopment analysis (DEA). The consensus in the performance benchmarking literature is that DEA is preferable in applications in which the frontier model cannot be expressed in algebraic form or does not have a known inefficiency distribution. The SFA method is preferable when certain classical assumptions are satisfied regarding the composite error terms, including the contributions from the inefficiency distribution and measurement errors. It has also been claimed that DEA has a comparative advantage in cases involving relatively small measurement errors due to the conceptual treatment of the errors, while the complementary SFA method has the advantage when the measurement errors are relatively high. A more thorough assessment of the two frontier estimation models may help managers responsible for performance-based budgeting to make more informed decisions regarding the most appropriate method for their particular circumstances.

Several studies have addressed the comparative advantages of stochastic versus deterministic frontier estimation. Banker, Gadh, and Gorr (1993) and Banker, Charnes, Cooper, and Maindiratta (1987) compared the efficiency estimation accuracy of corrected ordinary least squares (COLS) and DEA using Monte Carlo methods. The results indicated that DEA outperformed COLS in most cases and that COLS failed to distinguish between the measurement error and inefficiency. However, both frontier models failed as the measurement errors became large for all of the experimental scenarios considered. These results contradicted the traditional view favoring the use of stochastic frontier models. The DEA method has been criticized previously for its neglect of measurement errors; Greene (1993) even suggested that econometricians have abandoned the deterministic frontier model because it does not consider measurement errors. Gong and Sickles (1992) demonstrated the superiority of the SFA approach using Monte Carlo analysis; however, their results were criticized because of their assumption that the efficiency of the firm remains constant over time. In reality, the efficiency of a firm varies over time due to a variety of exogenous and endogenous factors. Ruggiero (1999) conducted another Monte Carlo analysis in which previous comparative studies were extended to include more general experimental scenarios and discovered that the deterministic frontier model which ignores the impact of measurement error is not as limited as the main criticism against the deterministic models stated negatively, but rather outperformed the stochastic frontier analysis model from the average rank correlation perspective.

Mixed results have therefore been obtained concerning the comparative advantages of stochastic versus deterministic frontier estimation. Nonetheless, one consistent finding is that DEA remains attractive as a frontier model, especially when the measurement errors become large. The traditional criticism of DEA based solely on the conceptual treatment of the errors should therefore be reconsidered. Another consistent finding is that as the measurement errors increase, the accuracy of the performance measurement decreases in both models. However, very little comparative research has been performed to date on how mismeasurement by frontier estimation models impacts the capital budget allocation and degree of system optimization.

Several studies in the DEA literature have addressed resource allocation based on efficiency analysis using variants of the DEA method. The goal of these studies is to balance the desires of two management layers, a central management authority and a set of operating units, by allocating the available resources in an optimal fashion. The balance is achieved by adjusting the input and output in such a way that the efficiency of each operating unit is maintained (the desire of the operating units) while the total output of units is maximized (the desire of the central management). Korhonen and Syrjanen (2004) developed a formal interactive approach based on DEA and multiple-objective linear programming to identify the optimal allocation plan. In this approach, the units are assumed to be capable of modifying their production within a specified production possibility set. Yan, Wei, and Hao (2002) extended the "inverse" DEA method by introducing preference cone constraints to allow decision makers to incorporate their preferences into the resource allocation algorithm. Li and Cui (2008) investigated an "efficient-effective-equality" resource allocation framework consisting of a DEA-based method leveraging many existing resource allocation algorithms. However, these DEA-based resource allocation approaches assume that sector-level decision making units are able to modify their production plants in a timely manner following instructions from the central management. In practice, this sort of rapid production plan modification is only possible in service firms such as supermarket chains, banks, universities, hospitals and tourist agencies. In the manufacturing industry, for instance, plants generally require a long time to adjust to new production plans, and the time required for a particular unit can vary depending on its operating conditions.

DEA has many opportunities and challenges under the multi-criteria environment. Mehdiabadi, Rohani, and Amirabdollahiyan (2013) proposed a new approach to combine DEA and Order Preference by Similarity to Ideal Solution (TOPSIS) to rank various industries which is also a multiple criteria decision making problem. Das, Sarkar, and Ray (2013) extended the proposed approach by Mehdiabadi et al. (2013) into fuzzy AHP–DEA-TOPSIS methodology which is applicable to any multiple criteria decision making problem due to its generic nature. Makui and Momeni (2012) considered similarities between multi-criteria decision making and DEA and tried to interpret decision makers preferences in UTA-STAR method using the common set of weights (CSW) in DEA.

The purpose of this paper is to perform a Monte Carlo analysis of different frontier estimation models combined with different multi-period budgeting strategies and to provide a set of decision rules for selecting the budgeting strategy and benchmark model that are most appropriate for a specified set of circumstances. Artificial measurement errors are included in the analysis.

The remainder of the paper is organized as follows. In Section 2, a replication study is performed to ensure the reliability of previous comparative study results on stochastic versus deterministic frontier estimation. Section 3 describes the experimental design and introduces the budgeting strategies, scenario generation methods and time-varying efficiency model used in this paper. The results of the experiments are presented in Section 4. An analytical proof is provided for the fact that a peanut butter strategy outperforms a discriminative strategy in the presence of large measurement errors, while the discriminative strategy is preferred when the measurement errors are small. Section 5 concludes with a summary of the findings of this study and suggestions for future research directions.

2. Comparison of SFA and DEA

Previous findings from related studies indicate that DEA and SFA have comparative advantages in the cases of small and large measurement errors, respectively. However, the accuracy of both frontier models decreases as the measurement error increases. A replication study is performed in this section to assure the reliability of these findings and to raise concerns regarding the use of frontier estimation models for a performance-based budgeting system in the presence of measurement errors.

Assume that a large organization includes multiple individual producers and desires a systematic method for measuring and comparing the performance of the various producers in the organization, including cases in which the inputs and outputs of the individual producers have different scales. The organization may use the performance information to identify opportunities for greater efficiency and to increase the overall system throughput and competitiveness. For simplicity, it can be assumed that each producer employs the same production process and therefore has the same production function. In this study, the Cobb-Douglas function for the n-th producer in the organization is assumed to have one output y, and two inputs, x_1 and x_2, assuming constant returns of scale although our mathematical analysis is independent of the returns of scale:

$$y_n = 2x_{1n}^{0.4}x_{2n}^{0.6}. \quad (1)$$

This functional form is in accordance with the models of Aigner and Chu (1968) and Ruggiero (1999). The inputs were generated randomly from a uniform distribution on the interval from 5 to 15. To express the difference between the actual production and frontier more realistically, the terms u_n and v_n are introduced into the translog form of Eq. (1), where u_n represents the half-normally distributed $[|N(0,\sigma_u^2)|]$ inefficiency and v_n denotes the normally distributed $[N(0,\sigma_v^2)]$ measurement error (including the data collection/reporting error and all other effects that are not accounted for in the analysis). The stochastic production frontier extending Eq. (1) to be used throughout this paper is then as follows:

$$\ln y_n = \ln 2 + 0.4\ln x_{1n} + 0.6\ln x_{2n} - u_n + v_n. \quad (2)$$

Eq. (2) gives the actual production of producer n, while Eq. (1) represents the best practice for product n. To distinguish y_n in Eq. (1) from y_n in Eq. (2), y_n in Eq. (1) will henceforth be denoted by $\hat{y}_n$. In Eq. (2), a firm-specific estimate of the Farrell efficiency is given by e^{-u_n}. For example, when u_n is assumed to be half-normally distributed according to $|N(0, 0.25^2)|$, the mean inefficiency is $0.2\left(\approx 0.25\sqrt{2}/\sqrt{\pi}\right)$ and the efficiency of a firm with the mean inefficiency becomes $82\%(\approx e^{-0.2})$. The study in this section considered six inefficiency distributions in which the value of σ_u^2 was varied between $0.1, 0.15, 0.2, 0.25, 0.3$, and 0.35. Seven measurement error distributions were considered with σ_v^2 taking values of $0.05, 0.1, 0.15, 0.2, 0.25, 0.3$, and 0.35. With six inefficiency distributions and seven measurement error distributions, a total of 42 experimental combinations were generated. In this short replication study, the sample size was restricted to 100. The "Benchmarking" package in R (Bogetoft & Otto, 2010) is used to implement the SFA and DEA.

The results of the Monte Carlo analysis are shown in Tables 1–6. This study uses two different performance metrics to evaluate the performance of SFA and DEA. The Spearman rank correlation is used by following Gong and Sickles (1992) and Ruggiero (1999) and also, MAD (mean absolute deviations) is used by following Banker et al. (1993). Therefore, Table 1 reports the average rank correlations between the true and SFA-estimated efficiencies. Tables 2 and 3 report the average rank correlations between the true and DEA-estimated efficiencies and between the SFA and DEA estimates, respectively. Similarly, Table 4 reports the MAD of true minus SFA-estimated efficiencies. Tables 5 and 6 report the MAD of true minus DEA-estimated efficiencies and SFA minus DEA-estimated efficiencies, respectively. All results are based on 100 replications. Three interesting results were obtained in this brief replication study.

Table 1
Rank correlation between the true efficiency and the SFA estimate.

	SFA vs. true (σ_v)						
	0.05	0.1	0.15	0.2	0.25	0.3	0.35
$\sigma_u = 0.1$	0.67	0.45	0.34	0.24	0.16	0.18	0.15
$\sigma_u = 0.15$	0.77	0.59	0.44	0.33	0.24	0.26	0.19
$\sigma_u = 0.2$	0.84	0.66	0.56	0.41	0.39	0.32	0.26
$\sigma_u = 0.25$	0.86	0.75	0.61	0.52	0.46	0.36	0.35
$\sigma_u = 0.3$	0.88	0.79	0.66	0.58	0.50	0.44	0.39
$\sigma_u = 0.35$	0.90	0.81	0.71	0.61	0.55	0.48	0.44

Table 2
Rank correlation between the true efficiency and the DEA estimate.

	DEA vs. true (σ_v)						
	0.05	0.1	0.15	0.2	0.25	0.3	0.35
$\sigma_u = 0.1$	0.52	0.34	0.25	0.19	0.12	0.17	0.11
$\sigma_u = 0.15$	0.60	0.44	0.31	0.26	0.19	0.20	0.15
$\sigma_u = 0.2$	0.65	0.50	0.43	0.34	0.30	0.25	0.23
$\sigma_u = 0.25$	0.66	0.60	0.46	0.38	0.38	0.26	0.30
$\sigma_u = 0.3$	0.69	0.64	0.51	0.46	0.37	0.35	0.32
$\sigma_u = 0.35$	0.73	0.63	0.55	0.49	0.45	0.39	0.33

Table 3
Rank correlation between the efficiencies estimated using SFA and DEA.

	SFA vs. DEA (σ_v)						
	0.05	0.1	0.15	0.2	0.25	0.3	0.35
$\sigma_u = 0.1$	0.69	0.70	0.69	0.73	0.71	0.72	0.71
$\sigma_u = 0.15$	0.73	0.73	0.70	0.71	0.72	0.74	0.72
$\sigma_u = 0.2$	0.74	0.72	0.72	0.71	0.72	0.75	0.72
$\sigma_u = 0.25$	0.73	0.74	0.73	0.73	0.75	0.72	0.73
$\sigma_u = 0.3$	0.75	0.75	0.75	0.75	0.72	0.72	0.75
$\sigma_u = 0.35$	0.77	0.76	0.75	0.76	0.73	0.73	0.73

Table 4
MAD of true minus SFA-estimated efficiencies.

	True vs. SFA (σ_v)						
	0.05	0.1	0.15	0.2	0.25	0.3	0.35
$\sigma_u = 0.1$	0.036	0.058	0.084	0.113	0.139	0.166	0.214
$\sigma_u = 0.15$	0.038	0.063	0.090	0.104	0.134	0.171	0.223
$\sigma_u = 0.2$	0.043	0.064	0.091	0.125	0.158	0.178	0.204
$\sigma_u = 0.25$	0.046	0.068	0.098	0.125	0.159	0.158	0.224
$\sigma_u = 0.3$	0.049	0.072	0.101	0.134	0.140	0.181	0.204
$\sigma_u = 0.35$	0.048	0.071	0.095	0.122	0.163	0.164	0.213

Table 5
MAD of true minus DEA-estimated efficiencies.

	True vs. DEA (σ_v)						
	0.05	0.1	0.15	0.2	0.25	0.3	0.35
$\sigma_u = 0.1$	0.041	0.071	0.098	0.129	0.152	0.175	0.187
$\sigma_u = 0.15$	0.047	0.073	0.100	0.127	0.151	0.173	0.188
$\sigma_u = 0.2$	0.052	0.077	0.102	0.131	0.154	0.176	0.188
$\sigma_u = 0.25$	0.057	0.077	0.104	0.130	0.151	0.176	0.196
$\sigma_u = 0.3$	0.062	0.081	0.105	0.131	0.154	0.179	0.190
$\sigma_u = 0.35$	0.068	0.087	0.107	0.132	0.151	0.175	0.190

Table 6
MAD of SFA minus DEA estimated efficiencies.

	SFA vs. DEA (σ_v)						
	0.05	0.1	0.15	0.2	0.25	0.3	0.35
$\sigma_u = 0.1$	0.035	0.056	0.081	0.100	0.127	0.164	0.186
$\sigma_u = 0.15$	0.040	0.063	0.085	0.098	0.123	0.156	0.192
$\sigma_u = 0.2$	0.047	0.064	0.082	0.116	0.145	0.158	0.178
$\sigma_u = 0.25$	0.055	0.067	0.091	0.114	0.146	0.143	0.203
$\sigma_u = 0.3$	0.059	0.071	0.094	0.130	0.138	0.167	0.188
$\sigma_u = 0.35$	0.066	0.079	0.097	0.119	0.146	0.153	0.200

First, SFA outperforms DEA regardless of the values adopted for the inefficiency and measurement errors (see Tables 1 and 2) when the rank correlation is used. Note that the higher correlation means the higher accuracy in the tables. Meanwhile, when MAD is used, the results show that the MAD of true minus SFA is smaller than that of DEA in the presence of small measurement errors ($\sigma_v \leqslant 0.15$), but the result is reversed for large measurement errors ($\sigma_v \geqslant 0.35$) (see Tables 4 and 5). In the middle range of measurement errors ($0.2 \leqslant \sigma_v \leq 0.3$), the MAD-based performances are mixed with two methods competing neck and neck. Since the smaller MAD is the more accurate, the results implies that SFA-estimated efficiency score is closer to the true efficiency than DEA in case that small measurement errors are present, but the result is reversed with large measurement errors. From this observations, it is found that the efficiency scores of SFA and DEA are sensitive to the change of magnitude of measurement error, leading to the plausibility of incurring severe mismeasurement. In the light of comparison with Banker et al. (1993), this results are consistent with theirs when the measurement errors is large. It also should be noted that since what they compared with DEA is not SFA but COLS[2] and furthermore SFA is considered more advanced than COLS in terms of separating noise from inefficiency, their results suggesting that DEA outperforms COLS make sense.

Second, the SFA and DEA estimates are highly correlated (staying around 0.7) in terms of their rank order regardless of the inefficiency and random error variation (see Table 3). This high correlation between the SFA and DEA estimates implies that the ranks in SFA and DEA are likely to be consistent and less fluctuate for varying measurement error, carrying both good and bad news for modelers. The good news is the degree of consistency between the two models, demonstrating the feasibility and robustness of the model estimations (Agrell & Bogetoft, 2007). The bad news is that it is difficult to use two complementary models to detect problems such as outlier presence or dominance in DEA or type-II error occurrence in SFA. In detail, when the DEA frontier estimate is biased high because of outlier data lying beyond the true frontier, the DEA method erroneously extends the estimated frontier outward. If the SFA method can distinguish between the inefficiency and noise with sufficient accuracy, then this method can be used in a complementary fashion to detect the DEA outlier problem. Similarly, DEA can be used in a complementary manner to detect the type-II error in SFA when the SFA frontier line reduces to a standard linear regression line. On the contrary, the MAD of SFA minus DEA efficiency scores increases from 0.035 to values greater than 0.2 as the measurement errors increases (see Table 6). The increasing MAD of SFA minus DEA efficiency scores means that although the accuracy in terms of efficiency scores deteriorates in both SFA and DEA, the deterioration of SFA is much faster than DEA because SFA is more sensitive to the change.

Third, the accuracy of the performance rank order in both models deteriorates significantly as the measurement errors increase. This third finding raises questions regarding the use of frontier estimation models in the presence of measurement errors, especially when a performance-based budgeting system is designed to allocate budgets to individual producers based on their results (e.g., performance ranks, efficiency scores). The impact of mismeasurement on future system performance and throughput potentials is also uncertain. In subsequent sections, the impact of mismeasurement is investigated using Monte Carlo analysis, and decision rules for the selection of a proper set of budgeting strategy and benchmark model in various scenarios are explored, with the goal of minimizing the loss in future system throughput potential.

[2] COLS (corrected ordinary least squares) produces consistent estimates based on the second and third moments of the OLS residuals; the biased intercept of the model is "corrected" based on the expected value of the composite error.

3. Experimental design

This study investigates the impact of mismeasurement by frontier estimation models experimentally using Monte Carlo methods for different budget allocation strategies and frontier estimation models.

As a reminder, the target environment under consideration in this paper is an organization with one central decision-making unit operating multiple individual producers where, in particular, low-performed producers may be strongly motivated to improve their performance because they have higher potentials to further improvement with less efforts than the higher ranked producers. Indeed, the effect of organizational competence stimulates less efficient producers to be more self-administrated for improvement. Many previous studies (e.g., Benjaafar, Li, & Daskin, 2013) have shown that organizational competence alone can lead to substantial performance improvement without significant increases in cost in many cases. For example, plant managers who received low rank in their scorecards are likely to be strongly motivated to improve their performance through operational adjustment as an alternative to costly hardware investment such as relocating equipment, choosing optimal suppliers, reducing energy intensity by picking a lot of low-hanging fruit in potential energy savings, and changing modes of transportation. Thus, both allocated budgets and operation competence can motivate each producers. The answer to which effect is larger among the aforementioned two factors in magnitude depends on particular circumstances specific to the industry or application under consideration, such as whether the producers are belonging to a for-profit organization or not, what is the return rate of capital projects implemented by additionally allocated budget or what is the cost saving rate of operational adjustment projects initiated by operation competence.

Under such an organization with a one central decision making unit operating multiple producers, two strategies are considered for the budget allocation in this study:

- A peanut butter budgeting strategy: Given a limited budget, B, for a single period and a collection of N producers, the n-th producer receives a budget proportional to its best-practice production size, $B\hat{y}_n/\sum_{j=1}^{N}\hat{y}_j$.
- A potential-weighted budgeting strategy: In this strategy, a budget B is allocated to each producer in proportion to its potential for production improvement, which is calculated as $Br_n\hat{y}_n/\sum_{j=1}^{N}r_j\hat{y}_j$ where r_n denotes the efficiency rank of the n-th producer. Note that producers with higher ranks have less potential for further improvement because of the decreased gap between $\hat{y}_n$ and y_n.

The "peanut butter" assumption is tenable and a common practice in the industry or business that are not willing to take a risk involved in allocating their resources differentially according to their potentials to realize. This assumption is taken, especially when the industry or businesses has no means to measure their achievable best practice level accurately. In such a case, they do not know what is going to work for improving their performance and instead, need to be across all areas, spreading their resources too thin across everywhere.[3]

Meanwhile, the potential-weighted budgeting strategy is a kind of performance-based budgeting and in this study, in particular, budgets are allocated based on ranks. The use of efficiency scores (SFA) or slacks (DEA) is also possible in the competitive nature

[3] In the costing literature, this peanut butter strategy has been recognized as one of traditional overhead cost allocation (or budgeting) methods where overhead costs is allocated based on a single relationship to numerous cost centers. Some experts refer to this traditional cost allocation approach as peanut butter accounting.

but they may be more applicable to use for the solo situation. Even worse, the efficiency scores or slacks may be unstable due to being sensitive to the change of measurement error as shown in Section 2. In the case that the impact of measurement errors is severe, a performance-based budgeting leveraging on frontier models may perform poorly. A better method, instead, could be to put the producers into buckets that may be produced in a certain range of ranks or efficiency scores and then, allocate the budgets to each meaningful bucket. In fact, the rank-order approach proposed in the paper is one possible configuration of the bucket arrangement.[4] Finding the optimal bucket arrangement is beyond the scope of this paper but the authors are interested in extending the research to address this issue as part of future research. In the formula of potential-weighted budgeting strategy, this paper sets a higher target (quota) for the lower ranked producers along with a higher budget to help them achieve the goals because the lower ranked producers have higher potentials to further improvement while the producers who are already closer to their best practice level tend to reach a point of diminishing returns. We also remark that there are many ways allocating the budget based on the reverse order of performance for the potential-weighted budgeting strategy, which depends on the environment. In this work, a simple linear model is used for the numerical simulation purpose.

Each budgeting strategy and performance rank calculation, based on either the true, SFA, DEA or randomly-generated ranks, is repeated over the specified time period. This study employed a new time evolution mechanism in which the budgeting strategy and performance rank calculation method influence the improvement in production over time. The time evolution mechanism is discussed in detail in Section 3.2.

3.1. Scenarios

In each experimental scenario, the performance rank of the n-th producer where $n = 1, 2, \ldots, N$ is evaluated in one of two ways:

- The true performance rank of each producer is calculated based on the true awareness of $\hat{y}_n$ and the value of v_n.
- The performance rank of each producer is estimated based on the lack of full awareness of $\hat{y}_n$ and the value of v_n.

Full awareness of the true performance ranks is unrealistic because it is difficult to accurately evaluate $\hat{y}_n$ and v_n for each producer. Although $\hat{y}_n$ and v_n are unknown, their values can be estimated based on the observed data using frontier estimation models. The use of estimated performance ranks is therefore the more realistic case. The estimated case can be further divided into three performance rank calculation approaches, two based on the frontiers estimated using SFA and DEA, and one in which the ranks are generated randomly. In a random ranking, each of the $N!$ permutations of the performance rank sequence is equally likely; this approach can provide a lower bound of the estimation performance.

For each performance rank calculation approach, either a peanut butter or potential-weighted budgeting strategy is adopted. A total of eight experimental scenarios are thereby generated for the Monte Carlo analysis, consisting of four different performance rank calculation approaches (true, SFA, DEA, and randomly-generated ranks) and two budgeting strategies (peanut butter and potential-weighted). The symbols "P" and "W" are used to denote the peanut butter and potential-weighted budgeting strategies, respectively, while the symbols "True","SFA", "DEA" and "Rnd" are used to distinguish the various performance rank calculation approaches. For example, "P-SFA" represents the scenario in which the performance rank of each producer is determined by measuring the discrepancy between its actual efficiency and its SFA-estimated frontier and the peanut butter budgeting approach is used to assign a budget to each producer. Similarly, "W-DEA" indicates a scenario in which the performance rank of each producer is determined by measuring the discrepancy between its actual efficiency and its DEA-estimated frontier and the potential-weighted budgeting approach is used to assign a budget to each producer. The true information-based performance ranking approach is expected to outperform the other scenarios, while the approach based on randomly-generated ranks is expected to provide a lower bound on the performance because it does not utilize any of the prior knowledge that is available from the observations.

3.2. Time-varying production efficiency

Slowly or quickly, firms respond to the needs and concerns of their customers and shareholders and adopt initiatives to improve their production. The conventional view is that such initiatives can be implemented in two ways: (1) adjustment of external factors, for example, through capital investment (or equivalent penalties or incentives) (2) operational adjustment through internal effort. As an example of capital investment, a firm may spend money to replace inefficient equipment and facilities. Alternatively, the firm may implement operational adjustments such as relocating facilities and choosing optimal suppliers and modes of transportation. While the value of external adjustments such as capital investments is clear, the economic potential of better business practices or operational policies is often overlooked. Many studies have shown that operational adjustments alone can lead to substantial performance improvement without significant increases in cost in many cases (Benjaafar et al., 2013).

This paper assumes that if producers are isolated from one another, their production efficiency can be improved by two drivers: external factors such as capital investment (or equivalent penalties or incentives) and internal operational adjustment. Based on this assumption, this paper derives a mathematical model of the time-varying production efficiency. Efficiency benchmarking is believed to trigger operational adjustment efforts by inducing each producer to compare its inefficiency against that of other firms. For example, if the efficiency ranking information is provided, producers with lower ranks will strive to improve their production even without external incentives such as capital investment or penalties imposed by headquarters. In summary, the ranking information itself may result in improved efficiency through the initiation of operational adjustment to pertinent producers.

As mentioned in the previous section, each producer is assumed to employ the same production process, and the output y_n is therefore given by the Cobb-Douglas model of Eq. (2). This assumption is valid for each time period and can be formulated by considering the weighted ratio of the output and inputs, although both the input and output are increasing functions of time t in general. That is, y_n is globally normalized in time. Eq. (1) therefore provides the best practice, which is denoted by $\hat{y}_n$. Note that the measurement error in Eq. (2) is not considered in this model, as it is assumed to be distributed uniformly throughout the time period. Therefore, y_n can be reduced as a function of the inefficiency only, which may depend on time t:

$$y_n = y_n(u_n). \tag{3}$$

The inefficiency therefore depends on two time-dependent factors, the external factor and the operational adjustment factor (or efficiency rank), which are denoted by c_n and r_n, respectively:

$$u_n = u_n(c_n(t), r_n(t)). \tag{4}$$

[4] The number of all partitions, B_N called 'Bell numbers', of N number of producers can be generated in vast ways (indeed, $B_n = \frac{1}{e}\sum_{k=0}^{\infty}\frac{k^n}{k!}$) and the rank-order is simple but one possible partition.

In what follows, a mathematical model is driven for the terms in Eq. (4). With r_n fixed, u_n is a function of the cost, c_n, only. In reality, c_n is in certain finite domain, say $[0, c_{max}]$, and the corresponding inefficiency u_n should

$$0 < u_{min} \leqslant u_n \leqslant u_{max} < \infty$$

because we expect u_n is finite even if there is no external factor at all, that is,

$$u_n(0) = u_{max} < \infty, \quad u_n(c_{max}) = u_{min} > 0.$$

For a moment, we denote e^{-u_n} by f_n (i.e. $u_n = -\ln(f_n)$), which represents the ratio of y_n and $\hat{y}_n$. Then f_n is a function of c_n as well and for some constants f_{min} and f_{max},

$$0 < f_{min} \leqslant f_n \leqslant f_{max} < 1.$$

Now we extend the range of f_n to $[0, 1]$ and the range of u_n to $[0, \infty]$. To this end, we introduce a potential cost c, which can be obtained from translation and extension of the range of c_n so that $0 < c < \infty$ and understood as an opportunity cost. Under assumption that f_n depends on the size of producer, we normalize the potential cost using $\hat{y}_n$. Then, we have

$$e^{-u_n} = f_n\left(\frac{c}{\hat{y}_n}\right).$$

At this point, the potential cost is assumed to be proportional to the size of the producer, which may be represented by its best-practice production.

Although it is more realistic to assume that larger firms achieve better efficiencies than smaller firms in production, in this work, we assume further that this returns to scale effect is relatively small. Then there is an efficient potential function, f independent of producers, such that

$$e^{-u_n} = f\left(\frac{c}{\hat{y}_n}\right), \quad 0 < c < \infty.$$

From this definitions of f and c, clearly u_n goes to infinity as c approaches 0 and the corresponding product is 0. On the other hand, as c goes to infinity, there is no inefficiency and the production reaches the best-practice level. In our mathematical argument (see Section 4), no specific forms of f are required as long as f satisfies that

(a) f is positive and increasing function with a horizontal asymptote $y = 1$, and
(b) f is concave.

Condition (a) comes from the definition of f and Condition (b), concavity of f, is a natural assumption because as the firm approaches its best practice level, $\hat{y}_n$, it becomes more difficult to further improve its performance assessment.

There are many choices for the potential function, f. Here, the following simple rational function is used for numerical simulations:

$$f\left(\frac{c}{\hat{y}_n}\right) = \frac{a_1(c/\hat{y}_n) + a_2}{a_3(c/\hat{y}_b) + a_4},$$

where $a_i (i = 1, 2, 3, 4)$ are coefficients to be determined from the assumptions

$$f(0) = 0,$$

$$\lim_{c\to\infty} f\left(\frac{c}{\hat{y}_n}\right) = 1.$$

In order to determine the rational function f explicitly, we need one more equation. Considering the rate of convergence of function f, we define a critical cost, $c^* = c_n^*/\hat{y}_n$, at which the performance of the n-th producer yields 99% of its best-practice production, i.e.,

$$f(c^*) = 0.99.$$

From all of the stated conditions, it follows that

$$f\left(\frac{c}{\hat{y}_n}\right) = \frac{99c}{99c + c^*\hat{y}_n}.$$

Note that the potential cost c for a given production level y_n can be represented as

$$c = f^{-1}\left(\frac{y_n}{\hat{y}_n}\right)\hat{y}_n. \tag{5}$$

Now we consider the operational adjustment by introducing a rank function:

$$c = g(r)\hat{y}_n, \quad 0 \leqslant r \leqslant 1,$$

where c is a potential cost. Through the rank function g, the operational adjustment can be converted to a potential cost c using the benchmarking methodology. As the efficiency rank is assumed to be the only information released, the operational adjustment r is a function of r_n, the rank of the n-th unit. More precisely, r is defined as follows:

$$r = \frac{r_n - 1}{N - 1}.$$

Recall that N is the total number of units. For simplicity in the current discussion, it is assumed that

$$r_j = 1 \quad \text{if } r_j < r_n. \tag{6}$$

That is, $g(r)$ is the rank function assuming that the ranks of all superior units are the same. With this assumption, we can avoid the complexity that may be yielded from various situations such as all the ranks of superior units are different or only some of them are the same.

Obviously, $g(0) = 0$ and $g(1) = c_a^*$, which is the maximum potential cost achievable through the operational adjustment. Furthermore, it is a natural assumption that the effort of the unit increases exponentially as its rank decreases. By adopting the form

$$g(r) = a(e^{br} - 1), \quad a, b > 0,$$

the following function is obtained:

$$g(r) = c_a^* \frac{\rho^2}{1 - 2\rho}\left(\left(\frac{1-\rho}{\rho}\right)^{2r} - 1\right).$$

Here, $0 < \rho < 1/2$ is a parameter that represents the factional increment in the potential cost, which satisfies

$$g\left(\frac{1}{2}\right) = \rho c_a^*.$$

In the general case, without assuming (6), we define the normalized potential cost due to the operational adjustment, $c_a/\hat{y}_n$, as a mean of $g(\frac{r_n - j}{N-1})$ for $j = 1, 2, \ldots, r_n$;

$$\frac{c_a}{\hat{y}_n} = \frac{1}{r_n}\sum_{j=1}^{r_n} g\left(\frac{r_n - j}{N - 1}\right). \tag{7}$$

Note that this definition can be understood as an approximation of the integral of g over $[0, r]$.

It is now possible to derive a recursion relation for $y_n(t_i)$, the production of the n-th producer at $t = t_i$. Suppose that the rank of the n-th producer is r_n and an incentive (or capital investment) $c_n(t_i)$ is assigned to this producer. The total potential cost for the n-th producer then increases to

$$\left(f^{-1}\left(\frac{y_n(t_i)}{\hat{y}_n}\right) + \frac{1}{r_n}\sum_{j=1}^{r_n} g\left(\frac{r_n - j}{N-1}\right) + \frac{c_n(t_i)}{\hat{y}_n}\right)\hat{y}_n.$$

It therefore follows that

$$y_n(t_{i+1}) = f\left(f^{-1}\left(\frac{y_n(t_i)}{\hat{y}_n}\right) + \frac{1}{r_n}\sum_{j=1}^{r_n} g\left(\frac{r_n - j}{N-1}\right) + \frac{c_n(t_i)}{\hat{y}_n}\right)\hat{y}_n.$$

Note that the parameters ρ and c_a^* that appear in the rank function are related to the circumstances of the system and the assigned budget, B. A smaller value of ρ may represent a relatively aggressive environment in that units with lower ranks are expected to expend more effort compared to less competitive circumstances. As for the choice of c_a^*, 50% of the average of $c_n/\hat{y}_n$ is adopted in our experiments. The critical cost c^* represents the current state of inefficiency as well as the slope of the potential function f. For example, a sufficiently large value of c^* yields a certain initial stage, but the marginal potential cost is higher. In this case, a long time period may be required to reach the best-practice level. In our experiments, c^* is chosen so that y_n approaches the best-practice level after 10 periods.

4. Results

The results of the Monte Carlo analysis are summarized in Figs. 2–5. Each scenario has a time horizon consisting of 10 decision epochs in which the performance rank of each producer is measured using the specified rank calculation approach and the budget is allocated based on the specified budgeting strategy.

The budget B for a single period is assumed to be fixed at 1% of the total best-practice production by the producers, $1\% \times \sum_{i=1}^{N}\hat{y}_i$. The maximal potential cost c_a^* and ρ values in the operational adjustment are taken to be 0.005 and 1/3, respectively. A value of 0.005 is obtained from 50% of the mean of the assigned normalized budget. In addition, we take $c^* = 0.2$, which is computed from the average of e^{-u_n} and the total cost gained after 10 periods.

Each scenario was ranked based on the total of production by all of the producers over the time horizon for the specified scenario. The measured rank was used as a performance metric to evaluate the individual scenario.

Eight scenarios were used for the Monte Carlo analysis (two budgeting strategies and four performance rank calculation approaches), and three factors were varied (10 inefficiency distributions, 15 measurement error distributions and 9 sample sizes), producing a total of 10,800 cells; 25 replications were performed for each cell.

Several interesting results were obtained. First, the peanut butter strategy performs better than the potential-weighted strategy in improving the overall system throughput in the presence of relatively high measurement errors, but the result is reversed for small measurement errors. As illustrated in Fig. 1, the potential-weighted budgeting strategy with SFA (line with solid circles) outperforms the peanut butter strategies based on SFA (line with empty circles) and DEA (line with empty triangles) up to $\sigma_u^2 = 0.15$, but after that point, the results are reversed. These experimental results demonstrate that the potential-weighted strategy outperforms the peanut butter strategy when the ranking information is accurate, while the peanut butter strategy is preferable to the potential-weighted strategy in the presence of relatively high measurement errors. These simulation results can be understood qualitatively by considering two extreme cases: when the true ranking information is known and when the rank of each producer is generated randomly.

In the first case, in which the accurate ranking information is released, the comparison between the peanut butter and potential-weighted budgeting strategies can be generalized to

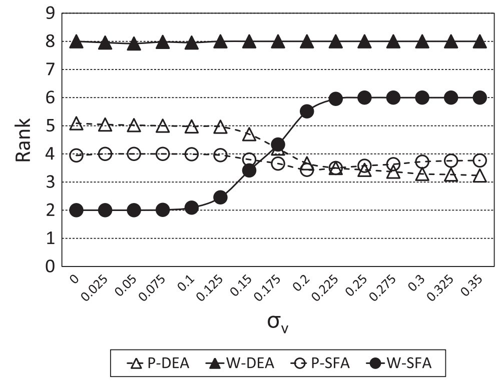

Fig. 1. Performance comparison (SFA vs. DEA).

the problem of seeking the optimal budgeting methods, which is related to the following optimization problem in which the impact of operation adjustments is ignored:

$$\max_{\mathbf{c}^+} F(\mathbf{c}_0 + \mathbf{c}^+)$$

$$s.t. \sum_{j=1}^{N} c_j^+ = C,$$

where

$$F(\mathbf{c}_0 + \mathbf{c}^+) = \sum_{j=1}^{N} f\left(\frac{c_{0j} + c_j^+}{\hat{y}_j}\right)\hat{y}_j.$$

$\mathbf{c}_0 = (c_{01}, c_{02}, \ldots, c_{0N})$ is the vector of the potential costs of the producers at the initial time, and $\mathbf{c}^+ = (c_1^+, c_2^+, \ldots, c_N^+)$ represents the costs gained during the entire period, where f is the potential function and $\hat{y}_n$ is the best-practice level of the n-th producer.

It can be shown using the method of Lagrange multipliers that F attains its maximum when all of the weighted potential costs, $(c_{0n} + c_n^+)/\hat{y}_n$ are the same. Indeed, from

$$\nabla_{\mathbf{c}^+} F(\mathbf{c}_0 + \mathbf{c}^+) + \lambda \nabla_{\mathbf{c}^+}\left(\sum_{j=1}^{N} c_j^+ - C\right) = 0,$$

we have

$$f'\left(\frac{c_{0n} + c_n^+}{\hat{y}_n}\right) + \lambda = 0, \quad \text{for all } n.$$

Because f is a one-to-one function, $(c_{0n} + c_n^+)/\hat{y}_n$ should be the same for all n. Let $(c_{0n} + c_n^+)/\hat{y}_n = \beta$. Then

$$c_n^+ = \beta\hat{y}_n - c_{0n}, \tag{8}$$

$$C = \beta\sum_{j=1}^{N}\hat{y}_j - \sum_{j=1}^{N} c_{0j}.$$

Substituting $\beta = \left(C + \sum_{j=1}^{N} c_{0j}\right) \Big/ \sum_{j=1}^{N}\hat{y}_j$ into (8) yields

$$c_n^+ = \frac{\hat{y}_n}{\sum_{j=1}^{N}\hat{y}_j}\left(C + \sum_{j=1}^{N} c_{0j}\right) - c_{0n} \quad \text{for all } n.$$

The total performance, $F(\mathbf{c})$, is therefore expected to increase as the variation of $\{c_j/\hat{y}_j\}$ decreases. For a given set of potentials $\{c_j/\hat{y}_j\}$ arranged in non-decreasing order, sufficiently small perturbations $\{\delta_j\}$ can be considered so that the order of $\{(c_j + \delta_j)/\hat{y}_j\}$ is preserved and

$$\sum_{i=j}^{N}\delta_j = 0,$$

$$\delta_j \geqslant 0 \quad 1 \leqslant j \leqslant J,$$
$$\delta_j < 0 \quad J+1 \leqslant j \leqslant N.$$

Clearly, the variance of $\{c_j/\hat{y}_j\}$ exceeds the variance of $\{(c_i+\delta_i)/\hat{y}_j\}$. As f' is non-increasing,

$$\begin{aligned}\boldsymbol{\delta}\cdot\nabla F(\mathbf{c}) &= \sum_{j=1}^{J}\delta_j f'\left(\tfrac{c_j}{\hat{y}_j}\right) - \sum_{j=J+1}^{N}(-\delta_j)f'\left(\tfrac{c_j}{\hat{y}_j}\right)\\ &\geqslant \sum_{j=1}^{J}\delta_j f'\left(\tfrac{c_J}{\hat{y}_J}\right) - \sum_{j=J+1}^{N}(-\delta_j)f'\left(\tfrac{c_J}{\hat{y}_J}\right)=0.\end{aligned}$$

$\delta=(\delta_1,\delta_2,\ldots,\delta_N)$ is therefore an ascent direction and for sufficiently small δ,

$$F(\mathbf{c})\leqslant F(\mathbf{c}+\boldsymbol{\delta}).$$

From this argument, one can deduce that the total performance with a potential-weighted strategy is higher than that with a peanut butter strategy. Indeed, for a given initial data or potentials $\mathbf{c_0}$, the variance of $\{(c_{0j}+c_j^+)/\hat{y}_j\}$ is fixed when a peanut butter approach is applied. In the case of potential-weighed strategy, however, $\{c_j^+/\hat{y}_j\}$ has a non-increasing order, which yields a smaller variance under the assumption of *reasonable weights that are not too large compared to the differences in the potentials*. In this case, superior total performance can be expected, as discussed above.

In real-world applications, however, the true ranks are often unknown; although the performance of each producer can be tracked using SFA or DEA, ranking information will be inaccurate due to a lack of information or large measurement errors. In this case, the aforementioned simulation results show that a peanut butter approach is superior to a potential-weighted strategy, as stated earlier. It is now assumed that the ranking information is entirely unknown or inaccurate, and each rank is therefore determined randomly.

To make the presentation explicit, a permutation $\gamma_k : i\to j$ is introduced, where i is the index of the producer and $\gamma_k(i)$ is the index of producer whose assigned ranking is the true ranking of the i-th producer in the k-th period. In the case in which the true ranking is known, γ_k is the identity transformation. In the case in which the rank is chosen at random, the expectation of $\gamma_k(i)$ over all possible permutations is constant for all i and k, and

$$E[c^+_{\gamma_k(i)}]=p_{ki},\quad 1\leqslant i\leqslant N, 1\leqslant k\leqslant M, \tag{9}$$

where $c^+_{\gamma_k(i)}$ is the cost assigned to i-th producer through a potential-weighed strategy based on its assigned rank and p_{ki} is the cost assigned to the i-th unit using the peanut butter approach for period k. Let I^+ and I^- be disjoint subsets of $\{1,2,\ldots,N\}$ such that

$$\begin{aligned}&c^+_{\gamma(i)}-p_i\geqslant 0,\quad \text{if } i\in I^+,\\ &c^+_{\gamma(i)}-p_i< 0,\quad \text{if } i\in I^-,\end{aligned}$$

where each entries of $\mathbf{c}^+=\{c^+_{\gamma(i)}\}$ and $\mathbf{p}=\{p_i\}$ represent the total assigned costs to i-th unit during whole period using the potential-weighted strategy based on $\gamma(i)=(\gamma_1(i),\gamma_2(i),\ldots,\gamma_M(i))$ and the peanut butter strategy, respectively. Then,

$$\begin{aligned}&F(\mathbf{c}_0+\mathbf{c}^+)-F(\mathbf{c}_0+\mathbf{p})\\ &=\sum_{i=1}^{N}\left[\int_{c_{0i}}^{c_{0i}+c^+_{\gamma(i)}} f'\left(\frac{x}{\hat{y}_i}\right)dx-\int_{c_{0i}}^{c_{0i}+p_i} f'\left(\frac{x}{\hat{y}_i}\right)dx\right]\\ &=\sum_{i\in I^+}\int_{c_{0i}+p_i}^{c_{0i}+c^+_{\gamma(i)}} f'\left(\frac{x}{\hat{y}_i}\right)dx+\sum_{i\in I^-}\int_{c_{0i}+p_i}^{c_{0i}+c^+_{\gamma(i)}} f'\left(\frac{x}{\hat{y}_i}\right)dx\\ &\leqslant \sum_{i\in I^+}(c^+_{\gamma(i)}-p_i)f'\left(\tfrac{c_{0i}+p_i}{\hat{y}_i}\right)\\ &\quad-\sum_{i\in I^-}\left(p_i-c^+_{\gamma(i)}\right)f'\left(\tfrac{c_{0i}+p_i}{\hat{y}_i}\right)\\ &=\sum_{i=1}^{N}(c^+_{\gamma(i)}-p_i)f'\left(\tfrac{c_{0i}+p_i}{\hat{y}_i}\right).\end{aligned}$$

The convexity of f is used in the above equations. By taking the expectation over all permutations, the following can be obtained from Eq. (9):

$$E[F(\mathbf{c}_0+\mathbf{c}^+)-F(\mathbf{c}_0+\mathbf{p})]\leqslant\sum_{i=1}^{N}(E[c^+_{\gamma(i)}]-p_i)f'\left(\frac{c_{0i}+p_i}{\hat{y}_i}\right)=0,$$

or

$$E[F(\mathbf{c}_0+\mathbf{c}^+)]\leqslant F(\mathbf{c}_0+\mathbf{p}).$$

Hence, we are able to conclude that the peanut butter approach is superior to any potential-weighted strategies in the mean sense as long as the ranking information is entirely unknown.

As mentioned earlier, the accuracy of SFA and DEA in determining the rank of each producer lies between the accuracy of the true and random rankings. Hence, one can conclude that the potential-weighted strategy performs better than a peanut butter strategy when the measurement error is sufficiently small; otherwise, a peanut butter strategy may yield better performance.

Note that a potential-weighed budgeting strategy coupled with DEA (line with solid triangles) exhibits the worst results, regardless of the variation in σ_u^2, as shown in Fig. 1. This poor performance occurs because the DEA-based ranking method generates many ties in its ranking of producers. The potential-weighed budget allocation with many ties in the ranking is likely to allocate $\mathbf{c}^+=(c_1^+,c_2^+,\ldots,c_N^+)$ in a relatively arbitrary manner to other cases and drive the overall system performance in a direction other than the optimal gradient direction.

The second interesting result is that the use of frontier estimation models is beneficial even for cases with relatively high measurement errors compared to using random ranking methods. Figs. 2–5 illustrate this result. Fig. 2 shows that the cases with accurate ranking information and the potential-weighted budget strategy (line with solid rectangles) always outperform the cases with SFA-estimated ranking information and the potential-weighted strategy (line with solid circles). Similarly, the cases with accurate ranking information and the peanut butter strategy (line with empty rectangles) always exhibit better results compared to the cases with SFA-estimated ranking information and the peanut butter strategy (line with empty circles). This result implies that the release of accurate ranking information always leads to greater performance improvement. Fig. 3 shows that the use of SFA-based estimation provides better results compared to those that can be obtained without using the model (equivalent to using a random ranking method), regardless of the budgeting strategy. This result is in accordance with the concept that the efforts of building estimation models are worthwhile when compared to DNA ("doing nothing analytical"). Similarly, Figs. 4 and 5 show that the DEA-based estimation approach falls between the cases using accurate ranking information and the cases using randomly-generated ranks.

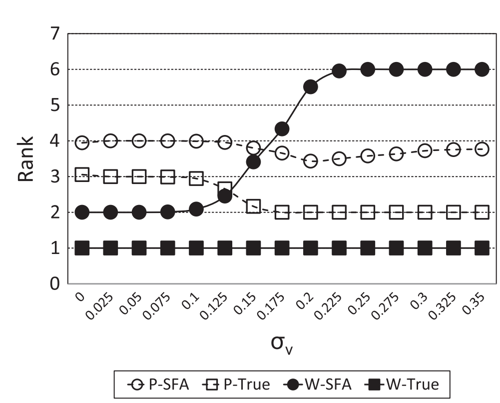

Fig. 2. Performance comparison (SFA vs. true ranking).

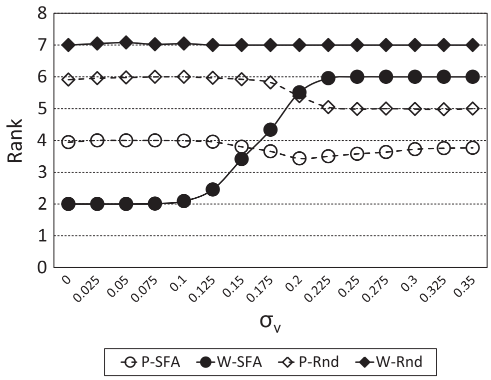

Fig. 3. Performance comparison (SFA vs. random ranking).

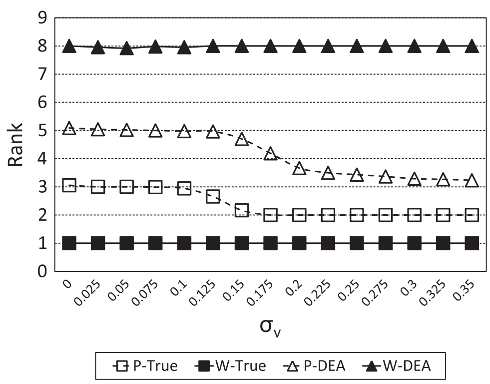

Fig. 4. Performance comparison (DEA vs. true ranking).

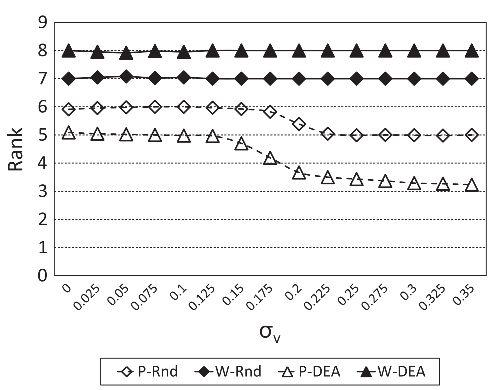

Fig. 5. Performance comparison (DEA vs. random ranking).

The third unexpected finding is that SFA performs better than DEA under peanut butter strategies, but the result is reversed as the measurement errors increase. In the case with high measurement errors, DEA tends to envelop points beyond the original production frontier due to positive measurement errors, resulting in an overestimated frontier. The DEA method also tends to generate many correlations in its ranking of the producers. The association between correlations in the ranking and an overestimated frontier may further increase the budget to certain producers. Fig. 1 confirms these speculations. The approach with SFA-estimated ranking information outperforms approaches with DEA-estimated ranking information, regardless of the budgeting strategy. As the measurement errors increase, however, the DEA approach becomes superior to approaches with SFA-estimated ranking information in cases using a peanut butter strategy. This result is in accordance with the first finding, which shows that peanut butter strategies perform better than potential-weighted strategies in the presence of relatively high measurement errors.

5. Conclusions

In this paper, a Monte Carlo analysis was performed to compare the overall system throughput in various experimental scenarios generated by pairing various multi-period budgeting strategies with two production frontier estimation models, SFA and DEA. Previously published studies only compared the accuracy of SFA and DEA efficiency estimation. This paper extended these studies by considering how mismeasurement by frontier estimation models impacts the capital budget allocation and final system performance. Decision rules are also provided for the selection of the most appropriate budgeting strategy and benchmark model for a particular set of circumstances.

One key conclusion from the analysis is that the selection of a proper budgeting strategy and benchmark model can lead to a substantial improvement in the system throughput. Three results that may be relevant for managers who are interested in performance-based budgeting are as follows: (1) a potential-weighted strategy performs better than a peanut butter strategy in improving the overall system throughput when the performance measurement errors are relatively small, but the result is reversed as the measurement error increases; (2) the use of frontier estimation models is beneficial even in cases with relatively high measurement errors compared to using randomly-generated ranks; and (3) SFA outperforms DEA regardless of the budgeting strategy for small measurement errors, but the result is reversed as the measurement errors increase.

Future studies should expand the scope of this work by further investigating optimal budget allocation strategies that may improve the overall system throughput much faster than the peanut butter or potential-weighted strategies proposed in this paper. As shown in Section 4, $\mathbf{c}^+ = (c_1^+, c_2^+, \ldots, c_N^+)$, the optimal budget for the entire period, should be chosen to lie between two vectors determined by the peanut butter and potential-weighted strategies, in such a way that $(c_{0n} + c_n^+)/\hat{y}_n$ is the same for all n. However, the challenge is that $\hat{y}_n$ varies over time for realistic environments due to constantly-evolving technology and, more importantly, the dynamics of competition between producers. In addition, the inefficiency and measurement error distributions are uncertain in real-world scenarios. As a result, the problem is more complicated compared to the case in which $\hat{y}_n$ is constant over time. Future studies should therefore extend the experiments in this paper to include the optimization of budget allocation by balancing the trade-off between peanut butter and potential-weighted strategies by evolving the current rank-order based bucketing method in the presence of time-varying best-practice frontier lines.

In addition, the authors admit that the assumption of using an identical measurement error distribution and an inefficiency distribution over time is not realistic in industry because it is conceivable that the distributions of measurement error and inefficiency may change dynamically over time. The dynamics behind time-varying inefficiency and measurement error distributions needs to be taken into account and accordingly, there is a need to develop a new dynamic SFA model to capture the dynamics. The authors are interested to extend the research to address this issue as part of future works.

References

Agrell, P., & Bogetoft, P. (2007). Development of benchmarking models for German electricity and gas distribution.

Aigner, D., & Chu, S. (1968). On estimating the industry production function. *American Economic Review, 58*, 826–839.

Banker, R., Charnes, A., Cooper, W., & Maindiratta, A. (1987). A comparison of data envelopment analysis and translog estimates of production frontiers using simulated observations from a known technology. In A. Dogramaci & R. Fare (Eds.), *Applications of modern production theory inefficiency and productivity*. Kenwer Academic Publishers.

Banker, R., Gadh, V., & Gorr, W. (1993). A Monte Carlo comparison of two production frontier estimation methods: Corrected ordinary least squares and data envelopment analysis. *European Journal of Operational Research, 67*, 332–343.

Benjaafar, S., Li, Y., & Daskin, M. (2013). Carbon footprint and the management of supply chains: Insights from simple models. *IEEE Transactions on Automation Science and Engineering, 10*(1), 99–116.

Bogetoft, P., & Otto, L. (2010). *Benchmarking with DEA, SFA, and R*. Springer.

Das, M., Sarkar, B., & Ray, S. (2013). A decision support framework for performance evaluation of indian technical institutions. *Decision Science Letters, 2*(4), 257–274.

Gong, B., & Sickles, R. (1992). Finite sample evidence on the performance of stochastic frontiers and data envelopment analysis using panel data. *Journal of Econometrics, 51*, 259–284.

Greene, W. (1993). The econometric approach to efficiency measurement. In H. Fried, C. Lovell, & S. Schmidt (Eds.), *The measurement of productive efficiency: Theory and applications*. New York: Oxford University Press.

Korhonen, P., & Syrjanen, M. (2004). Resource allocation based on efficiency analysis. *Management Science, 50*(8), 1134–1144.

Li, X., & Cui, J. (2008). A comprehensive dea approach for the resource allocation problem based on scale economies classification. *Journal of Systems Science and Complexity, 21*, 540–557.

Makui, A., & Momeni, M. (2012). Using csw weight's in utastar method. *Decision Science Letters, 1*(1), 39–46.

Mehdiabadi, A., Rohani, A., & Amirabdollahiyan, S. (2013). Ranking industries using a hybrid of dea-topsis. *Decision Science Letters, 2*(4), 251–256.

Ruggiero, J. (1999). Efficiency estimation and error decomposition in the stochastic frontier model: A Monte Carlo analysis. *European Journal of Operational Research, 115*, 555–563.

Yan, H., Wei, Q. L., & Hao, G. (2002). Dea models for resource reallocation and production input/output estimation. *European Journal of Operational Research, 136*(1), 19–31.