Xie, Q., L. L. Zhang, H. Shang, A. Emrouznejad, Y. Li (2021) Evaluating Performance of Super-Efficiency Models in Ranking Efficient Decision-Making Units based on Monte Carlo Simulations, Annals of Operations Research, https://doi.org/10.1007/s10479-021-04148-3

Evaluating Performance of Super-Efficiency Models in Ranking Efficient Decision-Making Units based on Monte Carlo Simulations

Qiwei Xie[1], **Linda L. Zhang**[*2,3], **Haichao Shang**[4], **Ali Emrouznejad**[5], **Yongjun Li**[6]

[1] School of Economics and Management, Beijing University of Technology, Beijing, 100124, China, qiwei.xie@ia.ac.cn

[2] IESEG School of Management, 1 parvis de la Défense, Socle de la Grande Arche, 92044 Paris-la Défense, France, l.zhang@ieseg.fr

[3] LEM-CNRS 9221, Université de Lille, Site Cité Scientifique Bâtiment SH2, 59655 Villeneuve D'Ascq, France

[4] School of Mathematics and Statistics, Hubei University, Wuhan, Hubei,430062, China, shanghc.hubu@gmail.com

[5]Aston Business School, Aston University, Birmingham B4 7ET, United Kingdom, a.emrouznejad@aston.ac.uk

[6] School of Management, University of Science and Technology of China, Hefei, Anhui 230026, China, lionli@ustc.edu.cn

Abstract

In response to the limitation of classical Data Envelopment Analysis (DEA) models, the super efficiency DEA models, including Andersen and Petersen (1993)'s model (hereafter; AP model) and Li et al. (2016)'s cooperative-game-based model (hereafter; L-L model), have been proposed to rank efficient decision-making units (DMUs). Although both models have been widely applied in practice, there is a paucity of research examining the performance of the two models in ranking efficient DMUs. Consequently, it is unclear how close the rankings obtained by the two models are to the "true" ones. Among the very few studies, Banker et al. (2015) pointed out that the ranking performance of the AP model is unsatisfactory; Li et al. (2016) and Hinojosa et al. (2017) demonstrated the L-L model's capability of ranking efficient DMUs without addressing the performance. This paper examines the ranking performance of the two super-efficiency models. In evaluating their performance, we carry out Monte Carlo simulations based on the well-known Cobb-Douglas production function and adopt Kendall rank correlation coefficient. Unlike Banker et al. (2015), we use the rankings obtained based on the two models and the "true" ones as the basis of performance evaluation in our simulations. Moreover, we consider several types of returns to scale (RS) and study the impact of changes of some parameters on the ranking performance. In view of the importance, we also carry out additional simulations to examine the influence of technical inefficiency on the two models' ranking performance. Based on the simulation results, we conclude: (1) Under different RS, the ranking performance of the two models remains the same when changing parameters, e.g., the distribution of input variables; (2) Under different RS, when technical inefficiency (compared with random noise) is more important, the two models have satisfactory performance by providing rankings that are close to, or the same as, the "true" ones; (3) The L-L model has better performance than the AP model and is more robust. (4) Under different RS, when technical inefficiency is less important, both models have unsatisfactory ranking performance; and (5) The relative importance of technical inefficiency plays an important role in ranking efficient DMUs.

Keywords: Monte Carlo; super efficiency, DEA, Ranking, Technical inefficiency, Returns to scale

1 Introduction

Based on the classical data envelopment analysis (DEA) models, such as the CCR model (Charnes et al., 1978) and the BCC model (Banker et al., 1984), the efficient scores calculated for efficient decision-making units (DMUs)[1] are all one. Consequently, the classical DEA models are not able to differentiate or rank efficient DMUs. Several ranking methods, such as cross-efficiency-based (Doyle and Green, 1994), common weight-based (Cook et al., 1990), Benchmark-based (Nicole Adler et al., 2002) and super-efficiency-based, have been proposed to rank efficient DMUs. these have been summarized in Appendix 1 along with their descriptions and limitations.

Here we describe in detail the super-efficiency-based ranking models, which we consider in this study. Andersen and Petersen (1993) introduced a super-efficiency DEA model (called the AP model in this study) to rank efficient DMUs. Unlike the classical DEA models, the AP model allows the super-efficiency of an efficient DMU to be larger than 1. This feature can find differences among efficient DMUs. Thanks to its simplicity and practicability, the AP model has been applied to evaluate DMUs in diverse industries, such as banks (Avkiran, 2011; Minh et al., 2013; Zimková, 2014), hospitals (O'Neill, 2005; Yawe, 2010; Du et al., 2014), and production-oriented enterprises (Düzakın and Düzakın, 2007; Zhang et al., 2012). It has also been employed to carry out energy evaluation (Li et al. ,2012; Li and Shi, 2014), player selection（Adhikari et al., 2020), and fixed cost allocation (Li et al., 2009; Lin and Chen, 2016; Dai et al., 2016).

Considering its popularity, Banker and Chang (2006) conducted simulations to evaluate the performance of the AP model in ranking efficient DMUs in a pioneering study. Based on the simulation results, they concluded that the AP model performs unsatisfactorily. A couple of years later, Banker et al. (2015) carried out an advanced simulation study to further test the AP model's ranking performance and drew the same conclusion. In both studies, the authors estimated the super efficiencies of efficient DMUs based on the AP model and calculated the "true" ones. They further calculated the corresponding Pearson and Spearman correlation coefficients. Because the correlations coefficients were negative, the authors concluded that the AP model's performance in ranking efficient DMUs is

[1] To be consistent with literature, in this study, efficient DMUs can be either strongly or weakly efficient.

not at all satisfactory. We doubt this conclusion because it is drawn based on the correlations of efficiencies, instead of rankings. First, the correlations of the "true" and super-efficiencies are not the same as these of the "true" rankings and the rankings obtained based on models. The results from (Xue and Harker, 2002) confirm this point. Second, even if the efficiencies are not correlated, the rankings from models might be close to the "true" ones. In relation to the above two points, the conclusion from Banker and Chang (2006) and Banker et al. (2015) may not be well grounded and needs to be further verified.

Li et al. (2016) proposed a cooperative-game-based super efficiency DEA model (called the L-L model) to rank efficient DMUs. According to this model, an efficient unit can have a score greater than one and a cooperative game with transferable utility can be defined to construct a general platform for efficient DMUs. The set of all efficient DMUs is regarded as a grand coalition with each efficient unit being a player. Based on the axioms of fairness, Shapley values are obtained for the cooperative game and sort the decision units. The literature (Li et al., 2016; Hinojosa et al., 2017) has demonstrated the capability of the L-L model in ranking efficient DMUs. In addition, Li et al. (2016) calculated the rankings of efficient DMUs based on the AP model, and found out differences between the rankings obtained based on their model (i.e., the L-L model) and the AP model. As explained by the authors, the differences were caused because the AP model evaluated efficient DMUs based on different efficient frontiers, whereas the L-L model employed one common platform in evaluation. However, the authors did not shed light on the performance of their model in ranking efficient DMUs. In other words, how close the rankings obtained based on their model are to the "true" ones is unknown. Similarly, Hinojosa et al. (2017) did not examine the ranking performance of the L-L model. In view of the above questionable conclusion from Banker and Chang (2006) and Banker et al. (2015) and the ranking differences between the L-L model and the AP model, we find it both interesting and relevant to investigate the ranking performance of two models.

To this end, in this study, we focus on the performance of the L-L model and the AP model in ranking efficient DMUs. Specifically, we carry out Monte Carlo simulations to examine their ranking performance[2]. Unlike Banker and Chang (2006) and Banker et al. (2015), we calculate and compare the correlations of rankings, instead of efficiencies, of efficient DMUs in the simulations. This is because the correlations of rankings can well explain if the rankings are close to one another, which can, in turn, describe the two models' ranking performance. Moreover, to include situations as comprehensive as possible, we consider several possible types of returns to scale (RS), including constant RS (CRS), decreasing RS (DRS), and increasing RS (IRS). (Note that RS is essential in calculating DMUs' efficiencies.) We also carry out simulations to examine if the changes of some parameters, e.g., the distribution of input variables, affect the ranking performance of the two models. Being an important parameter, technical inefficiency affects DMUs' efficiencies and further their rankings (Banker and Natarajan, 2008). Thus, we conduct additional simulations to examine the influence of the relative importance of technical inefficiency (to random noise) on the two models' ranking performance. Based on the simulation results, we draw several conclusions. First, under all three types of RS, parameter changes do not influence the ranking performance of the two models. Second, under all three types of RS, both models have satisfactory performance by providing rankings that are very close to, or the same as, the "true" ones when technical inefficiency is more important. Third, the L-L model performs better than the AP model does and is more robust. This is especially true when technical inefficiency is less important. Fourth, under all three types of RS, both models provide unsatisfactory performance when technical inefficiency is less important. Fifth, the relative importance of technical inefficiency plays an important role in ranking efficient DMUs. In accordance with the conclusions, we stress that (i) it is of paramount importance to determine suitable settings of technical inefficiency when ranking efficient DMUs and (ii) organizations are encouraged to apply the L-L model to rank their DMUs.

This study makes several contributions to the related literature. First, it identifies a critical issue in the conclusion from Banker and Chang (2006) and Banker et al. (2015). Second, it examines the ranking performance of the L-L model and the AP model using a new approach different from the one in (Banker and Chang, 2006; Banker et al., 2015). Third, with the simulation results, it offers several

[2] The simulation procedure in this study can also be used to rank inefficient DMUs.

conclusions and implications (see above).

The rest of this paper is organized as follows. In the next section, we provide a recap of the L-L model and the AP model. For details, the readers are referred to (Li et al., 2016) and (Andersen and Petersen, 1993). In Section 3, simulations and results are presented to evaluate the performance of the two models in ranking efficient DMUs. We present the additional simulations and results in Section 4 to shed light on the influence of technical inefficiency on the two models' ranking performance. In the last section, we summarize the results and contributions of our study and discuss possible avenues for future research.

2 Recap of two super-efficiency DEA models

Assume that there are n independent DMUs in a set N. A DMU: DMU_j $(j \in N = \{1,2,\dots,n\})$ consumes m inputs x_{ij} $(i \in M = \{1,2,\dots,m\})$ and produces h outputs y_{rj} $(r \in H = \{1,2,\dots,h\})$.

Input-oriented CCR model

For ease of reference, we first present the standard input-oriented CCR model as follows:

$$
\begin{aligned}
& E_f(N) = Min\theta \\
s.t. \quad & \sum_{j\in N} \lambda_j x_{ij} \le \theta x_{if},\ i \in M \\
& \sum_{j\in N} \lambda_j y_{rj} \ge y_{rf},\ r \in H \\
& \lambda_j \ge 0,\ \forall f, j \in N,\ \theta:\ free
\end{aligned}
\tag{1}
$$

Because the standard input-oriented CCR model cannot distinguish efficient DMUs, Andersen and Petersen (1993) proposed the AP model (see below).

AP model

$$E_f(N) = Min\theta \tag{2}$$

$\sum$

$$\sum_{\substack{j \in N \\ j \neq f}} \lambda_j y_{rj} \geq y_{rf},\ r \in H$$

$$\lambda_j \geq 0,\ \forall f, j \in N, j \neq f,\ \theta:\ free$$

If DMU_j is efficient, its efficiency calculated based on the AP model is no lower than 1; otherwise, its efficiency is lower than 1. This is consistent with the classical DEA models. However, there is a major difference between the AP model and the classical DEA models in evaluating DMUs. Unlike the classical DEA models, the AP model does not include a DMU under evaluation in the reference set. As a result, the super efficiencies of DMUs are calculated using different efficiency frontier surfaces, resulting in a multiplatform problem (Li et al., 2016). As shown in Figure 1 adapted from (Li et al., 2016), when evaluating B, the efficiency frontier surface is AB_1C, whereas the efficiency frontier surfaces are A_1BC and ABC_1 for evaluating A and C, respectively. The efficiencies of these DMUs are, thus, calculated based on different efficiency frontiers. This implies that the evaluation criteria for DMUs are different based on the AP model. (Note that the two axes: X_1 and X_2 in Figure 1 represent two inputs of the three DMUs: A, B, and C.)

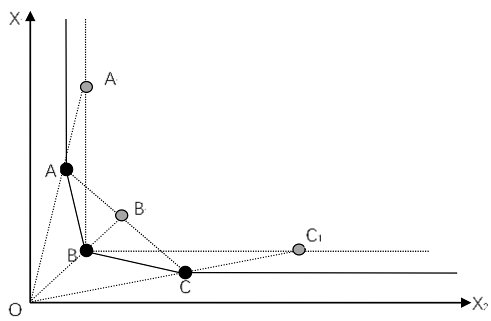

Figure 1: Different efficiency frontier surfaces for evaluating DMUs. Adapted from (Li et al., 2016)

Denote the set of efficient DMUs as $E^* = \{f: E_f(N) \geq 1\}$.

L-L model

With the definitions below, we summarize the L-L model.

Definition 1: The influence value of DMU d in E^* is called the efficiency change percentage (ECP), as defined below.

$$ECP_d(S) = \frac{E_d(N/S)}{E_d(N)} - 1 = E_d(N/S) - 1, \forall d \in E^*;\ \forall S \subset E^*. \tag{3}$$

N/S is the subset of N where S is removed. Based on the standard input-oriented CCR model, $E_d(N) = 1$. Because S is a subset of E^*, $E_d(N/S)$ is the super efficiency of DMU d based on the reference set N/S. Obviously, if $d \notin S, ECP_d(S) = 0$; if $d \in S$, $ECP_d(S) \geq 0$. Therefore, $ECP_d(S)$ indicates the influence degree of S on DMU d. An efficient DMU is called a player and E^* a grand coalition.

Definition 2: A common platform cooperative game with transferable utility (TU-game) is denoted as $< E^*, C >$:

(1) a finite set E^*: the set of players (i.e., efficient DMUs), and

(2) a function C assigned to each subset S (i.e., a coalition) of E^*: $\forall S \subset E^*$; a real number $C(S)$ exists. $C: 2^{E^*} \rightarrow R$ is denoted as the characteristic function, and $C(S)$ is the total payoff of coalition S.

Definition 3: Given a known S, the sum of $ECPs$ of all the DMUs in S is formulated below.

$$C(S) = \sum_{d \in S} ECP_d(S), \forall S \subseteq E^*. \tag{4}$$

$ECP_d(S)$ measures the influence of S on DMU d, whereas $C(S)$ measures its effect on the original efficiency frontier. $C(S)$ has the following property.

Property 1: Function C is super-additive: $C(S_1 + S_2) \geq C(S_1) + C(S_2)$ for arbitrary $S_1 \subset E^*$ and $S_2 \subset E^*$ as long as $S_1 \cap S_2 = \emptyset$.

As indicated by Property 1, Shapley value (Shapley, 1953) can be used to solve the cooperative game.

$$\varphi_k(C) = \sum_{k \in S, S \subset E^*} \frac{(s-1)!\,(p-s)!}{p!} [C(S) - C(S/\{k\})], \tag{5}$$

where p and s are the number of DMUs in E^* and S, respectively. S can be an arbitrary subset of E^*.

As highlighted in Definition 3, S cannot be equivalent to E^* when there are no inefficient DMUs in N. This is because the Shapley value method needs at least one inefficient DMU to compute $C(S)$. In case inefficient DMUs do not exist, a new DMU: denoted as E^- needs to be created with $X_i^- = \max_{k \in E^*}\{X_{ik}\}, i = 1, \dots, m$ and $Y_r^- = \max_{k \in E^*}\{Y_{rk}\}, r = 1, \dots, s$.

3 Ranking performance evaluation with Monte Carlo simulations

Simulation experiments based on Monte Carlo are carried out to evaluate the performance of the AP model and the L-L model in ranking efficient DMUs. In the simulations, the Cobb-Douglas production function is adopted; the basis for evaluation is the rankings of efficient DMUs obtained based on the AP model and the L-L model. Considering various combinations of the sample size and the number of inputs, we include diverse cases in the simulations.

Production function

Being widely accepted in literature, the Cobb-Douglas production function can well describe the technological relationship between inputs and outputs (Banker and Natarajan, 2008; Banker et al., 2004; Bjurek et al., 1990). In line with literature, the Cobb-Douglas production function (Eq. (6)) is adopted in this study.

$$Y = Ax_1^{\alpha_1} x_2^{\alpha_2} \dots x_m^{\alpha_m} e^{v-u}, \tag{6}$$

where Y is the output; $x_1, x_2, \dots, x_m$ are different inputs with *m* denoting the total number; $\alpha_1, \dots, \alpha_m$ are the elastic coefficients of *m* inputs.

The function captures the maximum output that can be obtained based on *m* inputs under the current technological conditions. In this study, $x_1, x_2, \dots, x_m$ independently follow a uniform distribution of $U(1, 4)$; $\alpha_1, \dots, \alpha_m$ independently follow a uniform distribution of $U(1/2m, 2/m)$; A is the comprehensive technology level and set to 1; u represents technical inefficiency and follows a nonnegative half-normal distribution: $|N[0, 5.06]|$; v denotes random noise and follows a normal distribution of $N(0, 0.36)$. The above settings are commonly adopted in literature (Chen and Delmas, 2012; Giraleas et al., 2012). When $v - u \geq 0$, a production process is efficient; otherwise, it is inefficient. e^{-u}, in fact, is the efficiency of a DMU under evaluation.

By following Chen and Delmas (2012), we transform Eq. (6) into Eq. (7), which is similar to Eq (12) in (Chen and Delmas, 2012).

$$\begin{bmatrix} \log y_1 \\ \vdots \\ \log y_n \end{bmatrix} = \begin{bmatrix} \log A + \alpha_1 \log x_1 + \cdots \alpha_m \log x_m \\ \vdots \\ \log A + \alpha_1 \log x_1 + \cdots \alpha_m \log x_m \end{bmatrix} + \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} - \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}, \tag{7}$$

where $y_1, \ldots, y_n$ are the output obtained for DMU_j; $x_1, \ldots, x_m$ are m inputs; $\alpha_1, \ldots, \alpha_m$ are the elastic coefficients of m inputs; $v_1, \ldots, v_n$ are random noise for DMU_j; $u_1, \ldots, u_n$ are technical inefficiency for DMU_j.

According to Coelli et al. (2005), the Cobb-Douglas production function exhibits three possible types of RS, including i) when $\alpha_1 + \alpha_2 + \cdots + \alpha_m = 1$, there are constant RS (CRS), ii) when $\alpha_1 + \alpha_2 + \cdots + \alpha_m > 1$, there are increasing RS (IRS), and iii) when $\alpha_1 + \alpha_2 + \cdots + \alpha_m < 1$, there are decreasing RS (DRS). To include different situations as more comprehensive as possible, the ranking performance of the two models is evaluated based on all these assumptions about RS. The simulation parameter settings are summarized in Table 1.

Table 1: Simulation parameter settings

Parameters	Value
Sample size	[10, 15, 20, 30]
Number of inputs	[4, 6, 8, 10]
Returns-to-scale parameters	$\sum \alpha_i = 1, \sum \alpha_i > 1, \sum \alpha_i < 1$
Distribution of input variables	$U\ (1, 4)$
Distribution of α_i	$U(1/2m, 2/m)$
Technical inefficiency distribution	$\mid N\ [0, 5.06] \mid$
Random noise distribution	$N\ (0, 0.36)$

We also conduct experiments to examine if the changes of some parameters affect the ranking performance of the two models. With the setting above, the parameters to be changed in these experiments include the distribution of input variables and the distribution of α_i (i.e., the elastic coefficients of input variables). These experiments are carried out under all three types of RS, including CRS (i.e., $\sum \alpha_i = 1$), IRS ($\sum \alpha_i > 1$), and DRS ($\sum \alpha_i < 1$). Table 2 provides the settings of these parameters, whilst the other parameter settings remain the same as above. As shown in the table, in the experiments, we i) change the distribution of input variables and ii) simultaneously change both distributions.

Table 2: Settings of parameter changes

Experiment	Parameter	Value
The first set of experiments	Distribution of input variables	$\lvert N\,[0,2.5] \rvert$
The second set of experiments	Distribution of input variables	$\lvert N\,[0,2.5] \rvert$
	Distribution of α_i	$\lvert N\,[1/2m, 2/m] \rvert$
The third set of experiments	Distribution of input variables	$U(1,6)$
The fourth set of experiments	Distribution of input variables	$U(1,6)$
	Distribution of α_i	$U(1/3m, 3/m)$

Simulation process and evaluation criteria

With the simulation parameter settings, random input values for DMUs are generated and the corresponding output values are obtained based on Eq. (7). Further, the "true" efficiencies of DMUs (i.e., e^{-u}) and the corresponding "true" rankings of efficient DMUs are obtained. R1 is used to represent the "true" rankings. Subsequently, with the same input and output values, the efficiencies and rankings of efficient DMUs are obtained based on the AP model and the L-L model. The rankings from the AP model and the L-L model are denoted as R2 and R3, respectively. R1, R2, and R3 are employed as the basis for evaluating the ranking performance of the two models.

Based on the Kendall rank correlation coefficient, the simulations compute the P-values of R1, R2, and R3. The smaller the P-value is, the closer the two rankings in consideration are. See Kendall and Gibbons (1990) for details about Kendall rank correlation. If the P-value is 0, the corresponding two rankings are identical. The P-value can, thus, indicate how close the rankings obtained based on the AP model and the L-L model are to the "true" ones. The P-value of R1 and R2 (or R3) is denoted as P_1 (or P_2). The test level is set to 0.05. If P_1 (or P_2) is larger than 0.05, there are statistically significant differences between R1 and R2 (or R3); otherwise, there is no statistically significant difference.

Let K be the total number of experiments in the simulation (K is 300 in this study). Denote the probability that there is no statistically significant difference between R1 and R2 as $\#\{P_1 \leq 0.05\}/K$, where $\#\{\mathrm{A}\}$ is the total number of event A in K experiments. Similarly, denote the probability that there is no statistically significant difference between R1 and R3 as $\#\{\mathrm{P}_2 \leq 0.05\}/K$. Consequently, the higher the probability is, the closer the corresponding ranking is to the "true" one. The values of P_1 and P_2 are compared. If $\mathrm{P}_2 < \mathrm{P}_1$, R3 is closer to R1 than R2 is. Accordingly, the L-L model has a better ranking performance than the AP model. If $\mathrm{P}_2 > \mathrm{P}_1$, R2 is closer to R1, indicating that the AP model performs better. If $\mathrm{P}_2 = \mathrm{P}_1$, two models exhibit the same ranking performance by generating identical rankings. In relation to the above three possible comparison results, three probabilities are defined. They are (i) the probability that the L-L model performs better: $\#\{\mathrm{P}_2 < \mathrm{P}_1\}/\mathrm{K}$, (ii) the probability that the two models have the same ranking performance: $\#\{\mathrm{P}_2 = \mathrm{P}_1\}/\mathrm{K}$, and (iii) the probability that the AP model performs better: $\#\{\mathrm{P}_2 > \mathrm{P}_1\}/\mathrm{K}$. The values of all five probabilities are calculated in the experiments.

Representing the total number of DMUs, the sample size *n* is set to 10, 15, 20, and 30 (see Table 1). For each sample size, we set the total number of inputs *m* to 4, 6, 8, and 10. Accordingly, there are sixteen different combinations of the sample size and the number of inputs. As per the three assumptions about RS, the experiments are classified into three groups. For the experiments involving CRS, we divide each of the randomly generated elastic coefficient values by the sum to get a new one. This step is to ensure the sum of the coefficient values to be 1, as required by CRS. Similarly, for the experiments involving IRS (or DRS), we divide the randomly generated coefficient values by the sum plus 1 (or by 0.8 times the sum). These two steps are to make the sum of the coefficient values less than 1 and greater than 1, respectively. In total, forty-eight cases are analyzed in the simulations.

Results and analysis

Tables 3-5 provide the results in terms of the values of the five probabilities under CRS/DRS/IRS. Further, the bar charts of the values of two probabilities $\#\{\mathrm{P}_1 \leq 0.05\}/K$ and $\#\{\mathrm{P}_2 \leq 0.05\}/K$ represented by p1 and p2 are developed in Figures 3-5. The bar charts can visualize how close the rankings obtained based on the two models are to the "true" ones. Under IRS, unlike the other cases, the case of n =10 and m=8 has a value of 0.9967, which is different from the other values in the first two rows (see Table 5 and Figure 5). This different value is caused by the randomness of the data.

As shown in Tables 3-5 and Figures 3-5, the results under three assumptions: CRS, DRS, and IRS have very similar characteristics, e.g., both the increasing and decreasing trends of the probability values. For illustrative simplicity, the analysis of the results under CRS are provided below. (Note that the same conclusions are obtained from analyzing the results under IRS and DRS.)

As shown in the first two rows of Table 3, no matter what values *n* and *m* take, all the probability values are 1. This indicates that there is no statistically significant difference between the two rankings and the "true" ones in all 300 experiments. As the rankings are the same as the "true" ones, we conclude that both models perform very well in ranking efficient DMUs under CRS. Moreover, both models perform well in ranking efficient DMUs when the number of DMUs (or inputs) increases.

Regarding the comparison of the ranking performance of the two models, we have the following observations. First, the values of $\#\{\mathrm{P}_2 \leq 0.05\}/K$ are no lower than those of $\#\{\mathrm{P}_1 \leq 0.05\}/K$ in all sixteen cases, as shown in the first two rows. Second, when m or *n* increases, the probability values pertaining to the two models remain unchanged (see p1 and p2 in Figure 2). Third, the sum of probability values in the third and fourth rows ($i.e., \#\{\mathrm{P}_2 \leq \mathrm{P}_1\}/K$) in each case is larger than 0.5 with the highest value being 1 (in all cases when n=20 and 30) and the lowest one 0.9533 (in the case when n=10 and m=8). (Note that the sum of probability values in the third and fourth rows in each case is the same as 1-the corresponding probability value in the fifth row.) Fourth, when n=20 and 30, no matter what values *m* takes, the rankings obtained based on the two models are identical, as shown by the 1s in the fourth row. With these observations, we conclude that under CRS, in general the L-L model performs better in ranking efficient DMUs than the AP model does.

We also obtain some other interesting observations. As shown in the fourth row, when the sample size *n* increases for the same number of inputs, the probability values increase in general. However, when the number of inputs *m* decreases for the same sample size, the probability values increase for most cases[3]. With these results, we conclude that when the sample size increases or the number of inputs decreases, the two models tend to generate same rankings. As proved in literature (Jenkins and Anderson, 2003; Pastor et al., 2002; Li and Liang, 2010), the number of efficient DMUs decreases in

[3] There are some exceptional cases, e.g., when n=10 and m=10.

the sample size or when the number of inputs decreases. The decreasing number of efficient DMUs may contribute to the higher probability of same rankings in our experiments. It is worth noting that when the sample size is large (n=20 or 30 in our simulations), the two models provide identical rankings.

To summarize, the simulation results are consistent across three RS assumptions. This indicates that the ranking performances of the two models are very similar. Additionally, when the number of DMUs increases or the number of inputs decreases, the probability that the two models generate same rankings increases under all three assumptions. At last, under all assumptions, both models provide satisfactory ranking results, and the L-L model performs more stable than the AP model.

Table 3: Simulation results under CRS

Probability	n=10				n=15				n=20				n=30			
	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10
$\#\{P_1 \le 0.05\}/K$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$\#\{P_2 \le 0.05\}/K$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$\#\{P_2 < P_1\}/K$	0.0267	0.0633	0.0600	0.0500	0.0133	0.0167	0.0067	0.0200	0	0	0	0	0	0	0	0
$\#\{P_2 = P_1\}/K$	0.9467	0.9000	0.8933	0.9100	0.9733	0.9800	0.9767	0.9700	1	1	1	1	1	1	1	1
$\#\{P_2 > P_1\}/K$	0.0267	0.0367	0.0467	0.0400	0.0133	0.0033	0.0167	0.0100	0	0	0	0	0	0	0	0

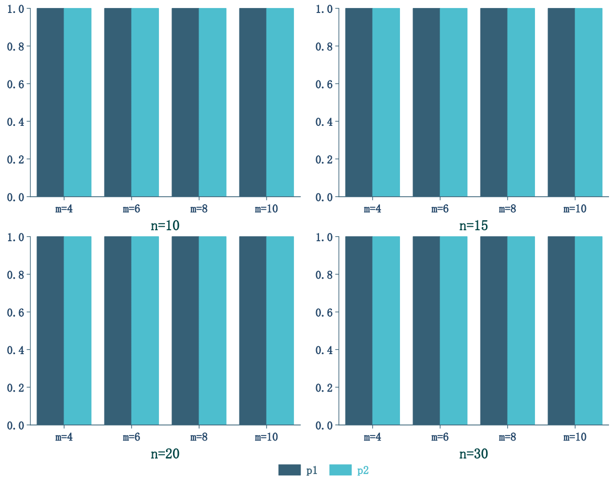

Figure 3: Probability bar charts under CRS

Table 4: Simulation results under DRS

Probability	n=10				n=15				n=20				n=30			
	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10
$\#\{P_1 \leq 0.05\}/K$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$\#\{P_2 \leq 0.05\}/K$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$\#\{P_2 < P_1\}/K$	0.0233	0.0333	0.0367	0.0700	0.0167	0.0067	0.0133	0.0100	0	0	0	0	0	0	0	0
$\#\{P_2 = P_1\}/K$	0.9600	0.9133	0.8933	0.9100	0.9833	0.9900	0.9800	0.9867	1	1	1	1	1	1	1	1
$\#\{P_2 > P_1\}/K$	0.0167	0.0533	0.0700	0.0200	0	0.0033	0.0067	0.0033	0	0	0	0	0	0	0	0

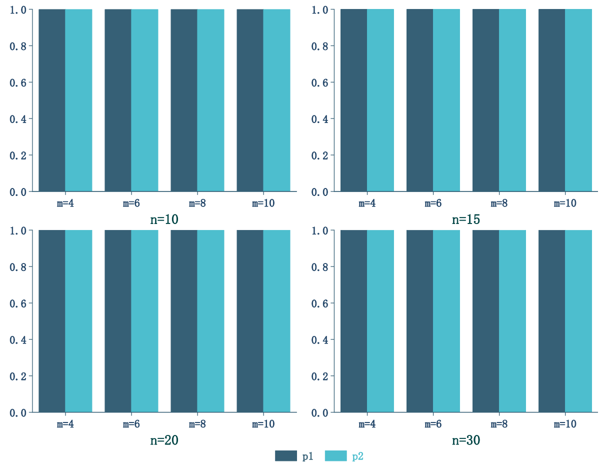

Figure 4: Probability bar charts under DRS

Regarding the results of the experiments where we changed the distribution of input variables and the distribution of α_i, they are consistent under all three types of RS. Thus, for illustrative simplicity, we provide the results under CRS in Tables 6-9 and Figures 6-9 below. (See the results under IRS and DRS in Appendix 2 and 3, respectively.) More specifically, Table 6 and Figure 6/Table 7 and Figure 7/Table 8 and Figure 8/Table 9 and Figure 9 correspond to the first/second/third/fourth set of experiments shown in Table 2. As shown in the first two rows of Table 6-9, the probability values are either 1 or very close to 1. This indicates that there is no statistically significant difference between the two rankings and the "true" ones in all 300 experiments.

Table 5: Simulation results under IRS

Probability	n=10				n=15				n=20				n=30			
	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10
$\#\{P_1 \leq 0.05\}/K$	1	1	0.9967	1	1	1	1	1	1	1	1	1	1	1	1	1
$\#\{P_2 \leq 0.05\}/K$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$\#\{P_2 < P_1\}/K$	0.0267	0.0300	0.0667	0.0500	0.0033	0.0167	0.0267	0.0067	0	0	0	0	0	0	0	0
$\#\{P_2 = P_1\}/K$	0.9633	0.9300	0.8867	0.9133	0.9933	0.9667	0.9667	0.9833	1	1	1	1	1	1	1	1
$\#\{P_2 > P_1\}/K$	0.0100	0.0400	0.0467	0.0367	0.0033	0.0167	0.0067	0.0100	0	0	0	0	0	0	0	0

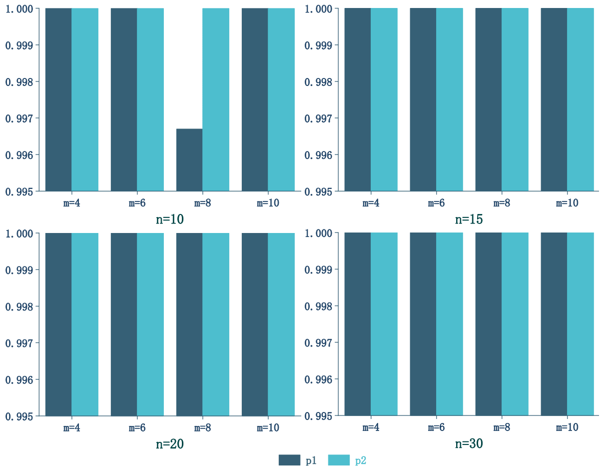

Figure 5: Probability bar charts under IRS

Consequently, we can conclude that both models have satisfactory ranking performance under CRS when the distributions of input variables and of α_i change. Moreover, by carefully examining the values in all the rows of the tables and figures, we found the same patterns as these in Tables 3-6 and Figures 3-6. Thus, we reach the final conclusion: When the distribution of input variables and the distribution of α_i change under CRS (as well as DRS and IRS), the two model's ranking performance does not change and remains the same as in the earlier simulations.

Table 6: Results of the first set of experiments (x~| N [0,2.5] |)

Probability	n=10				n=15				n=20				n=30			
	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10
$\#\{P_1 \leq 0.05\}/K$	0.9900	0.9967	0.9867	0.9900	1	1	1	1	1	1	1	1	1	1	1	1
$\#\{P_2 \leq 0.05\}/K$	0.9900	0.9967	0.9900	0.9967	1	1	1	1	1	1	1	1	1	1	1	1
$\#\{P_2 < P_1\}/K$	0.2000	0.3200	0.3667	0.4033	0.1733	0.2467	0.2700	0.3467	0.0100	0.0433	0.0767	0.0633	0	0	0	0
$\#\{P_2 = P_1\}/K$	0.6533	0.5467	0.4833	0.4433	0.7833	0.6833	0.6133	0.5867	0.9800	0.9533	0.9200	0.9200	1	1	1	1
$\#\{P_2 > P_1\}/K$	0.1467	0.1333	0.1500	0.1533	0.0433	0.0700	0.1167	0.0667	0.0100	0.0033	0.0033	0.0167	0	0	0	0

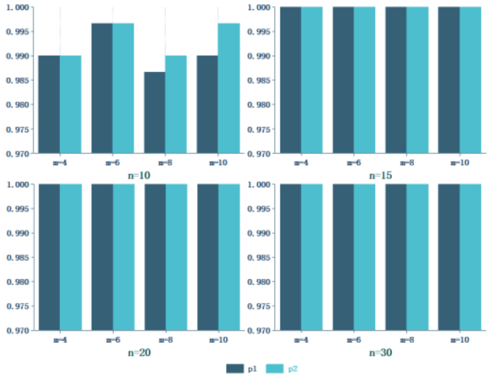

Figure 6: Probability bar charts corresponding to the first set of experiments

Table 7: Results of the second set of experiments (x~| N [0,2.5] | and α~| N $[1/2m, 2/m]$ |)

Probability	n=10				n=15				n=20				n=30			
	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10
$\#\{P_1 \leq 0.05\}/K$	0.9833	0.9900	0.9833	0.9800	1	1	1	1	1	1	1	1	1	1	1	1
$\#\{P_2 \leq 0.05\}/K$	0.9867	0.9900	0.9900	0.9867	1	1	1	1	1	1	1	1	1	1	1	1
$\#\{P_2 < P_1\}/K$	0.1633	0.3267	0.3667	0.3900	0.1433	0.2667	0.3200	0.3400	0.0267	0.0533	0.0800	0.1300	0	0	0	0.0033
$\#\{P_2 = P_1\}/K$	0.7133	0.5467	0.4733	0.4200	0.7800	0.6467	0.5667	0.5733	0.9700	0.9267	0.9000	0.8533	1	1	1	0.9967
$\#\{P_2 > P_1\}/K$	0.1233	0.1267	0.1600	0.1900	0.0767	0.0867	0.1133	0.0867	0.0033	0.0200	0.0200	0.0167	0	0	0	0

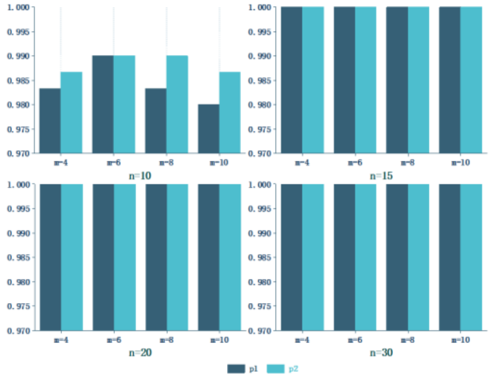

Figure 7: Probability bar charts corresponding to the second set of experiments

Table 8: Results of the third set of experiments ($x\sim U(1,6)$)

Probability	n=10				n=15				n=20				n=30			
	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10
$\#\{P_1 \le 0.05\}/K$	1	1	1	0.9967	1	1	1	1	1	1	1	1	1	1	1	1
$\#\{P_2 \le 0.05\}/K$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$\#\{P_2 < P_1\}/K$	0.0400	0.0800	0.0767	0.0767	0.0133	0.0233	0.0067	0.0233	0	0	0	0	0	0	0	0
$\#\{P_2 = P_1\}/K$	0.9200	0.8733	0.8600	0.8500	0.9867	0.9700	0.9833	0.9700	1	1	0.9967	1	1	1	1	1
$\#\{P_2 > P_1\}/K$	0.0400	0.0467	0.0633	0.0733	0.0000	0.0067	0.0100	0.0067	0	0	0.0033	0	0	0	0	0

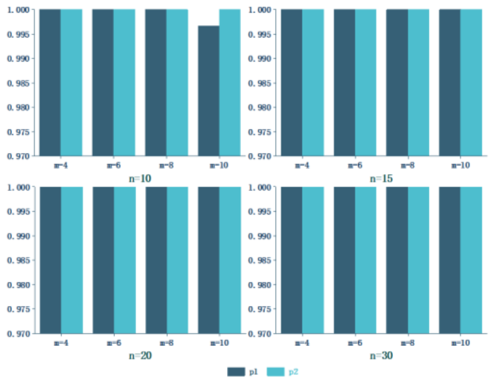

Figure 8: Probability bar charts corresponding to the third set of experiments

Table 9: Results of the fourth set of experiments ($x\sim U(1,6)$ and $\alpha\sim U(1/3m,3/m)$)

Probability	n=10				n=15				n=20				n=30			
	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10

$\#\{P_1 \leq 0.05\}/K$	1	1	1	0.9967	1	1	1	1	1	1	1	1	1	1	1	1
$\#\{P_2 \leq 0.05\}/K$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$\#\{P_2 < P_1\}/K$	0.0533	0.1000	0.0933	0.0800	0.0300	0.0367	0.0067	0.0333	0	0	0	0	0	0	0	0
$\#\{P_2 = P_1\}/K$	0.9067	0.8600	0.8400	0.8600	0.9567	0.9500	0.9700	0.9533	1	1	1	1	1	1	1	1
$\#\{P_2 > P_1\}/K$	0.0400	0.0400	0.0667	0.0600	0.0133	0.0133	0.0233	0.0133	0	0	0	0	0	0	0	0

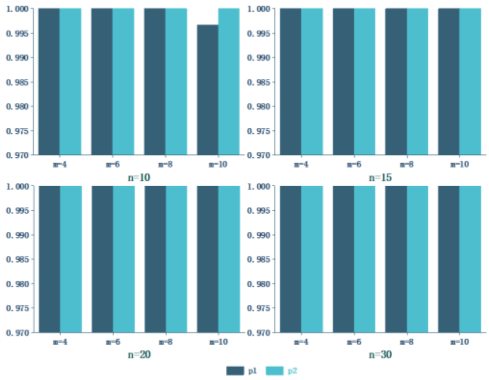

Figure 9: Probability bar corresponding to the fourth set of experiments

4 Simulation extension and result comparison

In the above simulations, being consistent with literature, we set $u \sim |N[0, 5.06]|$ and $v \sim N(0, 0.36)$. With these settings, the ratio of the variances of u and v is $\frac{\sigma_u^2}{\sigma_v^2} = \frac{5.06}{0.36} > 10$. This ratio, in fact, captures the relative importance of u to v. A high (or low) ratio indicates that u is more (or less) important (and accordingly, v is less (or more) important). When the ratio is high (or low), the two models may have satisfactory (or unsatisfactory) ranking performance. To find out the influence that the ratio may impose on the two models' ranking performance, additional simulations are carried out where the ratio values lie in two intervals: (1, 10) and (0, 1)[4]. More specifically, we conduct the simulations based on different settings of u while fixing v and the other parameters. Because the simulation results pertaining to different ratio values in each interval have the exactly same behaviors, we provide below the results and analysis related to i) $u \sim |N[0, 2.56]|$ and ii) $u \sim |N[0, 0.16]|$. For Case i), the ratio is $\frac{\sigma_u^2}{\sigma_v^2} = \frac{2.56}{0.36} \in (1,10)$, and for Case ii) $\frac{\sigma_u^2}{\sigma_v^2} = \frac{0.16}{0.36} \in (0, 1)$.

[4] The results and conclusions remain the same when different intervals, e.g., (0, 2), (2, 11), are used.

The simulation results are provided in Tables 10-12 and Figures 10-12 when $\frac{\sigma_u^2}{\sigma_v^2} = \frac{2.56}{0.36}$ and in Tables 13-15 and Figures 13-15 when $\frac{\sigma_u^2}{\sigma_v^2} = \frac{0.16}{0.36}$.

Table 10: Simulation results under CRS when $u \sim |N[0,2.56]|$

Probability	n=10				n=15				n=20				n=30			
	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10
$\#\{P_1 \leq 0.05\}/K$	0.9900	0.9833	0.9933	0.9900	0.9967	1	1	1	1	1	1	1	1	1	1	1
$\#\{P_2 \leq 0.05\}/K$	0.9867	0.9833	0.9933	0.9933	0.9967	1	1	1	1	1	1	1	1	1	1	1
$\#\{P_2 < P_1\}/K$	0.1000	0.1300	0.1167	0.1400	0.0667	0.1200	0.1233	0.1333	0.0067	0.0067	0.0200	0.0133	0	0	0	0
$\#\{P_2 = P_1\}/K$	0.8300	0.8000	0.7800	0.7733	0.8700	0.7967	0.8133	0.8000	0.9900	0.9833	0.9633	0.9800	1	1	1	1
$\#\{P_2 > P_1\}/K$	0.0700	0.0700	0.1033	0.0867	0.0633	0.0833	0.0633	0.0667	0.0033	0.0100	0.0167	0.0067	0	0	0	0

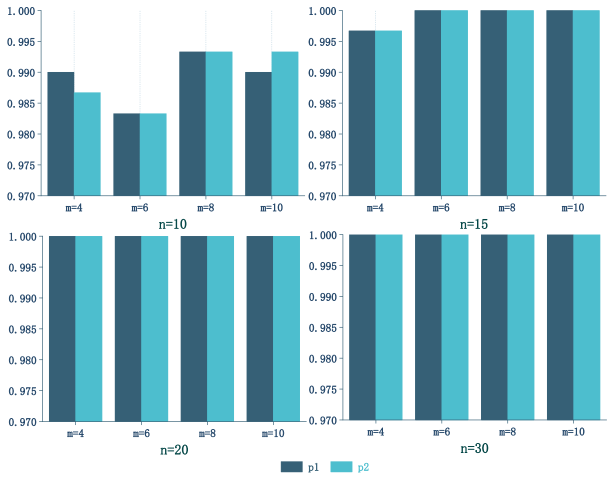

Figure 10: Probability bar charts under CRS when $u \sim |N[0,2.56]|$

Table 11: Simulation results under DRS when $u \sim |N[0,2.56]|$

Probability	n=10				n=15				n=20				n=30			
	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10
$\#\{P_1 \leq 0.05\}/K$	0.9900	0.9833	0.9867	0.9900	1	1	1	1	1	1	1	1	1	1	1	1
$\#\{P_2 \leq 0.05\}/K$	0.9900	0.9833	0.9967	0.9867	1	1	1	1	1	1	1	1	1	1	1	1
$\#\{P_2 < P_1\}/K$	0.0933	0.1267	0.1600	0.1500	0.0800	0.1233	0.1167	0.1700	0.0200	0.0067	0.0233	0.0067	0	0	0	0

$\#\{P_2 = P_1\}/K$	0.8500	0.8033	0.7367	0.7400	0.8500	0.8233	0.8033	0.7333	0.9700	0.9867	0.9600	0.9867	1	1	1	1
$\#\{P_2 > P_1\}/K$	0.0567	0.0700	0.1033	0.1100	0.0700	0.0533	0.0800	0.0967	0.0100	0.0067	0.0167	0.0067	0	0	0	0

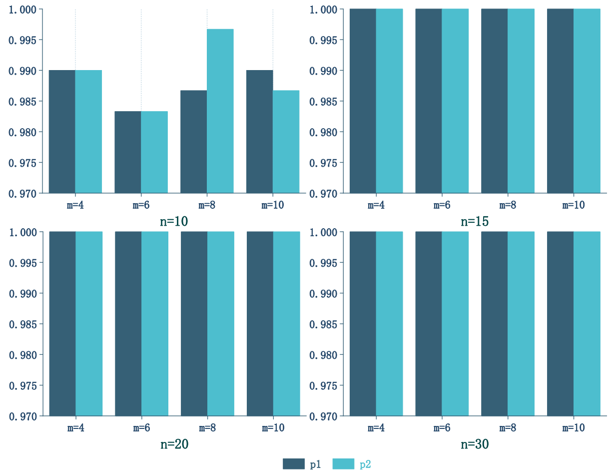

Figure 11: Probability bar charts under DRS when $u \sim |N[0,2.56]|$

Table 12: Simulation results under IRS when $u \sim |N[0,2.56]|$

Probability	n=10				n=15				n=20				n=30			
	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10
$\#\{P_1 \leq 0.05\}/K$	0.9967	0.9900	0.9967	0.9933	1	1	1	1	1	1	1	1	1	1	1	1
$\#\{P_2 \leq 0.05\}/K$	0.9967	0.9900	0.9933	0.9967	1	1	1	1	1	1	1	1	1	1	1	1
$\#\{P_2 < P_1\}/K$	0.0867	0.0933	0.1333	0.1233	0.0600	0.1000	0.1167	0.1100	0.0200	0.0133	0.0200	0.0233	0	0	0	0
$\#\{P_2 = P_1\}/K$	0.8633	0.8300	0.7767	0.7733	0.8900	0.8100	0.8267	0.7867	0.9767	0.9833	0.9600	0.9633	1	1	1	1
$\#\{P_2 > P_1\}/K$	0.0500	0.0767	0.0900	0.1433	0.0500	0.0900	0.0567	0.1033	0.0033	0.0033	0.0200	0.0133	0	0	0	0

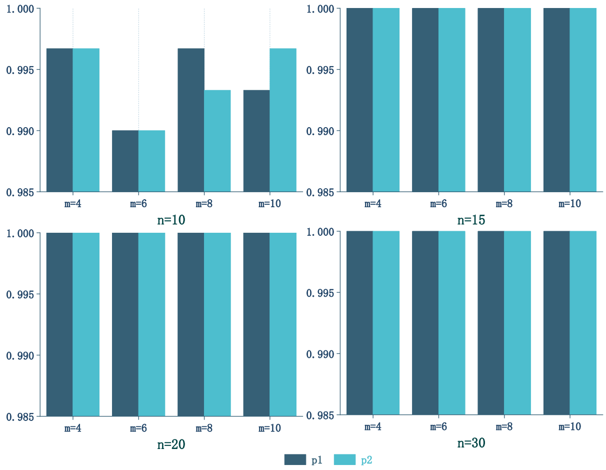

Figure 12: Probability bar charts under IRS when $u \sim |N[0, 2.56]|$

Table 13: Simulation results under CRS when $u \sim |N[0, 0.16]|$

Probability	n=10				n=15				n=20				n=30			
	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10
$\#\{P_1 \leq 0.05\}/K$	0.0833	0.0733	0.0900	0.1033	0.1067	0.1200	0.1000	0.1033	0.1167	0.1500	0.0933	0.1367	0.2267	0.2767	0.2167	0.1800
$\#\{P_2 \leq 0.05\}/K$	0.0900	0.0733	0.0833	0.1067	0.1033	0.1267	0.1000	0.1033	0.1200	0.1633	0.0967	0.1400	0.2200	0.2700	0.2267	0.1733
$\#\{P_2 < P_1\}/K$	0.2400	0.2533	0.2700	0.3733	0.3167	0.3833	0.3267	0.3533	0.2967	0.3633	0.4400	0.4667	0.3067	0.4133	0.4167	0.4467
$\#\{P_2 = P_1\}/K$	0.4933	0.3967	0.4033	0.3467	0.4667	0.3233	0.3300	0.2633	0.4667	0.3133	0.2033	0.1500	0.4300	0.2167	0.1867	0.1200
$\#\{P_2 > P_1\}/K$	0.2667	0.3500	0.3267	0.2800	0.2167	0.2933	0.3433	0.3833	0.2367	0.3233	0.3567	0.3833	0.2633	0.3700	0.3967	0.4333

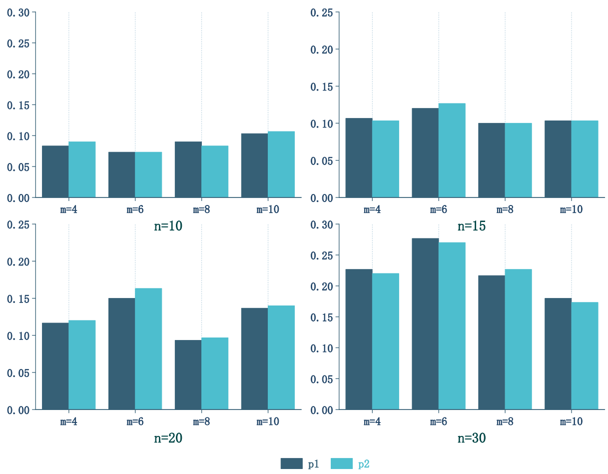

Figure 13: Probability bar charts under CRS when $u \sim |N[0, 0.16]|$

Table 14: Simulation results under DRS when $u \sim |N[0, 0.16]|$

Probability	n=10	n=15	n=20	n=30

	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10
$\#\{P_1 \le 0.05\}/K$	0.0833	0.0700	0.0833	0.0867	0.1000	0.0667	0.0933	0.1033	0.1133	0.1267	0.1000	0.1133	0.1967	0.2133	0.1900	0.2300
$\#\{P_2 \le 0.05\}/K$	0.0933	0.0867	0.0900	0.1033	0.1033	0.0667	0.1200	0.1233	0.1100	0.1200	0.1133	0.1067	0.1900	0.2067	0.1867	0.2300
$\#\{P_2 < P_1\}/K$	0.2633	0.3433	0.3433	0.3333	0.2933	0.3333	0.4100	0.4300	0.2667	0.3867	0.4100	0.4067	0.3433	0.4367	0.4500	0.4000
$\#\{P_2 = P_1\}/K$	0.5067	0.4000	0.3400	0.3667	0.4500	0.3600	0.2567	0.2167	0.4333	0.3067	0.2167	0.1767	0.3633	0.1933	0.1600	0.1700
$\#\{P_2 > P_1\}/K$	0.2300	0.2567	0.3167	0.3000	0.2567	0.3067	0.3333	0.3533	0.3000	0.3067	0.3733	0.4167	0.2933	0.3700	0.3900	0.4300

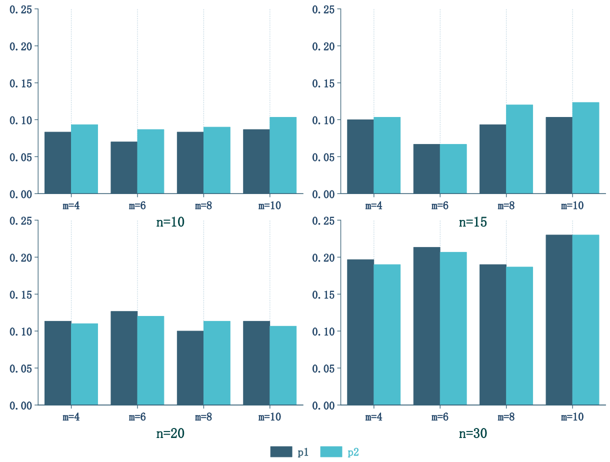

Figure 14: Probability bar charts under DRS when $u \sim |N[0, 0.16]|$

Table 15: Simulation results under IRS when $u \sim |N[0, 0.16]|$

Probability	n=10				n=15				n=20				n=30			
	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10
$\#\{P_1 \le 0.05\}/K$	0.0800	0.0800	0.1000	0.0500	0.1200	0.1233	0.1067	0.1367	0.1467	0.1233	0.1400	0.2167	0.1867	0.1867	0.2067	0.2433
$\#\{P_2 \le 0.05\}/K$	0.0767	0.0800	0.1067	0.0467	0.1300	0.1233	0.1100	0.1333	0.1467	0.1167	0.1300	0.1967	0.1967	0.1900	0.2100	0.2467
$\#\{P_2 < P_1\}/K$	0.2233	0.3133	0.2933	0.3633	0.2400	0.3533	0.4200	0.4367	0.2600	0.3567	0.3267	0.3933	0.3300	0.4333	0.4433	0.4100
$\#\{P_2 = P_1\}/K$	0.5767	0.3600	0.4100	0.2867	0.5000	0.3400	0.2667	0.2533	0.4800	0.3133	0.2433	0.2200	0.3733	0.2233	0.1400	0.1900
$\#\{P_2 > P_1\}/K$	0.2000	0.3267	0.2967	0.3500	0.2600	0.3067	0.3133	0.3100	0.2600	0.3300	0.4300	0.3867	0.2967	0.3433	0.4167	0.4000

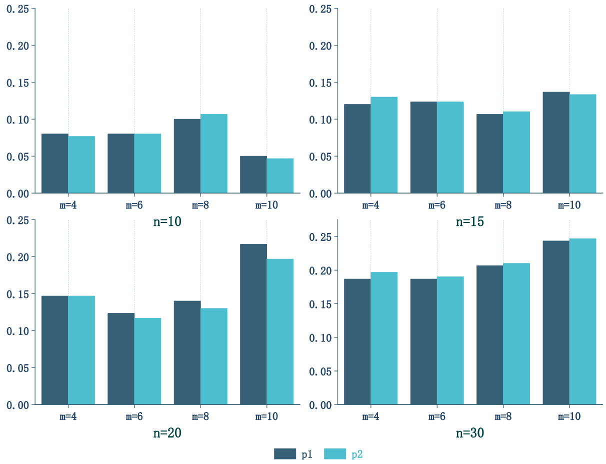

Figure 15: Probability bar charts under IRS when $u \sim |N[0, 0.16]|$

When $\frac{\sigma_u^2}{\sigma_v^2} = \frac{2.56}{0.36} \in (1,10)$, the results under CRS, DRS and IRS exhibit very similar characteristics. Thus, the analysis of the results under CRS and their comparison with those of the initial simulations are provided below. As shown in the first two rows of Table 10, the probability values are very close to 1, and in more than a half of the cases they are 1. This indicates that the two models perform very well in ranking efficient DMUs when $\frac{\sigma_u^2}{\sigma_v^2} \in (1,10)$. By comparing Table 10 with Table 3 where the probability values of the first two rows are all 1, we conclude that the ranking performance of the two models increases in the ratio of the variances of u and v. Based on the first two rows, we can also conclude that the ranking performance of both models increases with the sample size n. Despite some irregular cases, e.g., when n=10 and m=6 for the L-L model, when n=10 and m=10 for the AP model, the two models generally present improved ranking performance when the number of inputs m increases. After examining the probability values in the rest three rows of Table 10, we conclude that in general the L-L model performs better than the AP model. Same as in the initial simulations, when the sample size n is large, the two models exhibit the same performance in ranking efficient DMUs.

When $\frac{\sigma_u^2}{\sigma_v^2} = \frac{0.16}{0.36} \in (0,1)$ indicating that technical inefficiency is less important, the results (in Tables 13-15 and Figures 13-15) have some characteristics different with the earlier ones. Accordingly, this leads to some new conclusions. The results under CRS, DRS, and IRS exhibit very similar characteristics. Similarly, for illustrative simplicity, the analysis and conclusions based on Table 13 under CRS are presented below.

As shown in the first two rows, most probability values are close to 0 with the highest one being 0.2767 for the AP model (when n=30 and m=6) and the lowest 0.0733 for both models (when n=10 and m=6). Such low values indicate that the rankings in consideration are inconsistent with the "true" ones. We, therefore, conclude that both models have unsatisfactory ranking performance when technical inefficiency is less important. Despite the unsatisfactory performance, the probability values present an increasing trend in the sample size *n* (or the number of inputs *m*). Thus, we draw a similar conclusion to the earlier one; that is, when $\frac{\sigma_u^2}{\sigma_v^2} \in (0,1)$, the ranking performance of the two models increases in the sample size (or the number of inputs). Based on the probability values in the last row (or the sum of the corresponding values in the third and fourth rows) and the comparison of the first two rows, we conclude that under CRS, the L-L model performs better than the AP model. Moreover, as shown in the third row, with the increase of the sample size or the number of inputs, the probability values increase, indicating that the L-L model performs better and better. Based on the above analysis, we point out that the L-L model is more robust than the AP model when technical inefficiency is less important.

Finally, we summarize all the simulation results and conclusions. Under CRS, DRS, and IRS, when $\frac{\sigma_u^2}{\sigma_v^2} > 10$ and $\frac{\sigma_u^2}{\sigma_v^2} \in (1,10)$ where technical inefficiency is more important, the two models generate rankings, which are close to, or the same as, the "true" ones. However, when $\frac{\sigma_u^2}{\sigma_v^2} \in (0,1)$ where technical inefficiency is less important, they generate inconsistent rankings. We, therefore, conclude: (i) Under all types of RS, when technical inefficiency is more important, both models have satisfactory ranking performance, and provide rankings, which are close to, or the same as, the "true" ones; and (ii) Under all types of RS, when technical inefficiency is less important, both models perform unsatisfactorily in ranking efficient DMUs. Moreover, we conclude that in general, the L-L model performs better than the AP model. This is especially true when technical inefficiency is less important. This conclusion implies that organizations should try to employ the L-L model to rank their efficient DMUs. Last, we stress that the relative importance of technical inefficiency plays an important role in ranking efficient DMUs. In practice, it is, thus, of paramount importance to determine suitable settings of u when either of the two models is applied to rank efficient DMUs.

5 Conclusions

In literature, two super-efficiency DEA models proposed by Andersen and Petersen (1993) and Li et al. (2016) have been widely applied to differentiate or rank efficient DMUs. In this study, we called these two models the AP model and the L-L model, respectively. Monte Carlo simulations to evaluate the performance of the two models in ranking efficient DMUs were carried out. Some of these simulations were carried out to check if the changes of parameters (e.g., the distribution of input variables) affect the ranking performance of the two models. In view of its importance, we conducted additional simulations to examine the influence of technical inefficiency's variability on the two models' ranking performance. In the simulations, to include situations as comprehensive as possible, we considered three types of RS, including CRS, DRS, and IRS. The simulation results highlighted: (i) Under all three types of RS, the ranking performance of the two models remains the same when changing the parameters. (ii) Under the three types of RS, when technical inefficiency is more important, both models have satisfactory performance; (iii) The L-L model performs better than the AP model does and is more robust. This is especially true when technical inefficiency is less important; (iv) Under the three types of RS, both models provide unsatisfactory ranking performance when technical inefficiency is less important; and (v) The relative importance of technical inefficiency plays an important role in ranking efficient DMUs. With the results, we suggested that (a) Organizations might apply the L-L model in evaluating their DMUs and (b) Suitable settings of technical inefficiency are required to evaluate DMUs.

In the simulations, we assumed that the distribution of random noise remained the same. As random noise may affect DMUs' efficiencies, researchers interested could determine the suitable settings of random noise for optimizing the ranking performance of the L-L model. Furthermore, in this study, we used a maximum of 30 DMUs in the simulation, even though with 30 DMUs the calculation is very timely, thus, it would be interesting to design algorithms/mechanisms so that the number of DMUs can be increased in larger experiments while incurring acceptable simulation time. This is about reducing computational complexity. Additionally, one possible avenue for future research is to adopt other types of production functions in simulations and to examine if same results can be obtained.

References

[1] Adhikari, A., Majumdar, A., Gupta, G. & Bisi, A. (2020). An innovative super-efficiency data envelopment analysis, semi-variance, and Shannon-entropy-based methodology for player selection: Evidence from cricket. Annals of Operations Research, 284(1): 1-32.

[2] Adler N., Friedman L., Sinuany-Stern Z. (2002). Review of ranking methods in the data envelopment analysis context. European Journal of Operational Research, 140(2):249-265.

[3] Andersen, P., & Petersen, N.C. (1993). A procedure for ranking efficient units in data envelopment analysis. Management science, 39(10), 1261-1264.

[4] Avkiran, N.K. (2011). Association of DEA super-efficiency estimates with financial ratios: Investigating the case for Chinese banks. Omega, 39(3), 323-334.

[5] Banker, R.D., Charnes, A., & Cooper, W.W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. Management science, 30(9), 1078-1092.

[6] Banker, R.D., & Chang, H. (2006). The super-efficiency procedure for outlier identification, not for ranking efficient units. European Journal of Operational Research, 175(2), 1311-1320.

[7] Banker, R.D., Chang, H., & Zheng, Z. (2015). On the use of super-efficiency procedures for ranking efficient units and identifying outliers. Annals of Operations Research, 116(1-4), 1-15.

[8] Banker, R.D., Cooper, W.W., Seiford, L.M., Thrall, R.M., & Zhu, J. (2004). Returns to scale in different DEA models. European Journal of Operational Research, 154(2), 345-362.

[9] Banker, R.D., & Natarajan, R. (2008). Evaluating contextual variables affecting productivity using data envelopment analysis. Operations research, 56(1), 48-58.

[10] Bjurek, H., Hjalmarsson, L., & Forsund, F.R. (1990). Deterministic parametric and nonparametric estimation of efficiency in service production: A comparison. Journal of Econometrics, 46(1-2), 213-227.

[11] Charnes, A., Cooper, W.W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. European journal of operational research, 2(6), 429-444.

[12] Chen, C.M. & Delmas, M.A. (2012). Measuring eco-inefficiency: A new frontier approach. Operations Research, 60(5):1064-1079.

[13] Giraleas, D., Emrouznejad, A., & Thanassoulis, E. (2012). Productivity change using growth accounting and frontier-based approaches-Evidence from a Monte Carlo analysis. European Journal of Operational Research, 2012, 222(3):673---683.

[14] Coelli, T.J., Rao, D.S. P., O'Donnell, C.J., & Battese, G.E. (2005). An Introduction to Efficiency and Productivity Analysis (Springer, New York).

[15] Cook W. D., Roll, Y., & Kazakov, A. (1990). DEA model for measuring the relative efficiencies of highway maintenance patrols. Information Systems & Operational Research, 28(2):113-124.

[16] Dai, Q., Li, Y., & Liang, L. (2016). Allocating fixed costs with considering the return to scale: A DEA approach. Journal of Systems Science and Complexity, 29(5):1320-1341.

[17] Doyle J. & Green R. (1994). Efficiency and Cross-efficiency in DEA: Derivations, Meanings and Uses. Journal of the Operational Research Society, 45(5):567-578.

[18] Düzakın, E., & Düzakın, H. (2007). Measuring the performance of manufacturing firms with super slacks based model of data envelopment analysis: An application of 500 major industrial enterprises in Turkey. European Journal of Operational Research, 182(3), 1412-1432.

[19] Du, J., Wang, J., Chen, Y., Chou, S.Y., & Zhu, J. (2014). Incorporating health outcomes in Pennsylvania hospital efficiency: An additive super-efficiency DEA approach. Annals of Operations Research, 221(1):161-172.

[20] Hinojosa M. A., Lozano S., Borrero D. V., et al. (2017). Ranking efficient DMUs using cooperative game theory. Expert Systems with Applications, 80(9):273-283.

[21] Jenkins, L. & Anderson, M. (2003). Multivariate statistical approach to reducing the number of variables in data envelopment analysis. European Journal of Operational Research, 147, 51-61.

[22] Kendall, M.G. & Gibbons, J.D. (1990). Rank correlation methods. (Oxford University Press, New York).

[23] Liang L., Jie W., Cook W. D., et al. (2018). Alternative secondary goals in DEA cross-efficiency evaluation. International Journal of Production Economics, 113(2):1025-1030.

[24] Li, H. & Shi, J.F. (2014). Energy efficiency analysis on Chinese industrial sectors: An improved Super-SBM model with undesirable outputs. Journal of Cleaner Production, 65:97-107.

[25] Li, J.Y. & Cheng, Y.X. (2012). Total-factor energy efficiency analysis of Hebei province based on super efficiency DEA. Industrial Engineering Journal, 15(1):87–92.

[26] Li, L. & Liang, L. (2010). A Shapley value index on the importance of variables in DEA models. Expert Systems with Applications, 37(9), 6287-6292.

[27] Lin, R. & Chen, Z. (2016). Fixed input allocation methods based on super CCR efficiency invariance and practical feasibility. Applied Mathematical Modelling, 40:5377-5392.

[28] Li, Y., Yang, F., Liang, L., & Hua, Z. (2009). Allocating the fixed cost as a complement of other cost inputs: A DEA approach. European Journal of Operational Research, 197(1), 389-401.

[29] Li, Y., Xie, J., Wang, M., & Liang, L. (2016). Super efficiency evaluation using a common platform on a cooperative game. European Journal of Operational Research, 255(3), 884-892.

[30] Minh, N.K., Long, G.T., Hung, N.V. (2013). Efficiency and super-efficiency of commercial banks in Vietnam: Performances and determinants. Asia Pacific Journal of Operational Research, 30(1):1250047-1-19.

[31] O'Neill, L. (2005). Methods for understanding super-efficient data envelopment analysis results with an application to hospital inpatient surgery. Health Care Management Science, 8(4):291-298.

[32] Pastor, J.T., Ruiz, J.L., & Sirventm I. (2002). A statistical test for nested radial DEA models. Operations Research 50, 728-735.

[33] Shapley, L.S. (1953). A value for N-person games. Contribution to the Theory of Games, Vol. 2. Princeton University Press, Princeton, NJ, 307-317.

[34] Sexton, T. R., Silkman R. H., & Hogan A. J. (2010). Data envelopment analysis: Critique and extensions. New Directions for Evaluation, 1986(32):73-105.

[35] Xue, M. & Harker, P.T. (2002). Note: Ranking DMUs with infeasible super-Efficiency DEA models. Management Science, 48(5):705-710.

[36] Yawe, B. (2010). Hospital performance evaluation in Uganda: A super-efficiency Data Envelope Analysis model. Zambia Social Science Journal.

[37] Zhang, Q.Z., He, F., & Zhao, X. (2012). Analysis of Chinese energy efficiency of iron and steel industry based on super-efficiency DEA. Soft Science, 29:1320-1341.

[38] Zimková, E. (2014). Technical efficiency and super-efficiency of the banking sector in Slovakia. Procedia Economics and Finance, 12:780-787.

Appendix 1: Ranking methods for sorting DMUs and their limitations

The cross-efficiency sorting method was first proposed by Doyle & Green (1994). In the DEA background, this method is to sort the DMUs based on a cross-efficiency matrix. The cross-efficiency matrix was originally developed by Sexton et al. (1986), who started the theme of DEA ranking. Doyle and Green (1994) pointed out that decision makers do not always have a reasonable mechanism for ranking units and, thus, recommended the cross-efficiency matrix as a hierarchical unit. The cross-efficiency ranking method uses the optimal weights evaluated by n linear programs (LPs) to calculate the efficiency score of each DMU n times. The results of all DEA cross efficiency scores can be organized as the cross-efficiency matrix. The cross-efficiency soring method allows each DMU to selfishly choose an optimal set of input and output weights and defines the average of a DMU's efficiencies based on the optimal weights as its cross efficiency. However, cross-efficiency scores are generally not unique and depend on the alternative optimal solutions to the LPs used (Liang and Wu et al., 2008b). The efficiency obtained by this method is, thus, not unique.

The Common weight-based method is also used for ranking DMUs proposed in (Cook et al., 1990). It attempts to find a common set of weights to calculate efficiencies for all DMUs and further ranks them based on their efficiencies. Compared with the cross-efficiency method, this method evaluates the DMUs based on a common platform, i.e., the set of common weights. In this regard, the efficiency scores of units are comparable. The problem with this ranking method is that it is very difficult to find an inclusion principle to choose a common weight set. Consequently, the different principles adopted result in different common weight sets and grades between DMUs.

Nicole Adler et al. (2002) discussed the Benchmark ranking method, which sorts DMUs in two stages. In the first stage, the efficient units are ranked by simply counting the number of times they appear in the reference sets of inefficient units. The inefficient units are then ranked, in the second stage, by counting the number of DMUs that needs to be removed from the analysis before they are considered efficient. However, a complete ranking cannot be assured because many DMUs may receive the same ranking score. Another problem with this ranking method is that a DMU is highly ranked if it is chosen as a useful target for many other DMUs.

Appendix 2: Results of simulation where parameters are changed under DRS

Table 16: Results of the fifth set of experiments (x~| N [0,2.5] |)

Probability	n=10				n=15				n=20				n=30			
	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10
$\#\{P_1 \leq 0.05\}/K$	0.9900	0.9867	0.9767	0.9900	1	1	1	1	1	1	1	1	1	1	1	1
$\#\{P_2 \leq 0.05\}/K$	0.9900	0.9967	0.9900	0.9967	1	1	1	1	1	1	1	1	1	1	1	1
$\#\{P_2 < P_1\}/K$	0.2533	0.3633	0.4033	0.3833	0.1500	0.2867	0.3433	0.3467	0.0433	0.0700	0.0900	0.0633	0	0	0.0033	0
$\#\{P_2 = P_1\}/K$	0.6267	0.5433	0.4867	0.5033	0.7800	0.6167	0.5767	0.5867	0.9400	0.9133	0.9067	0.9200	1	1	0.9967	1
$\#\{P_2 > P_1\}/K$	0.1200	0.0933	0.1100	0.1133	0.0700	0.0967	0.0800	0.0667	0.0167	0.0167	0.0033	0.0167	0	0	0	0

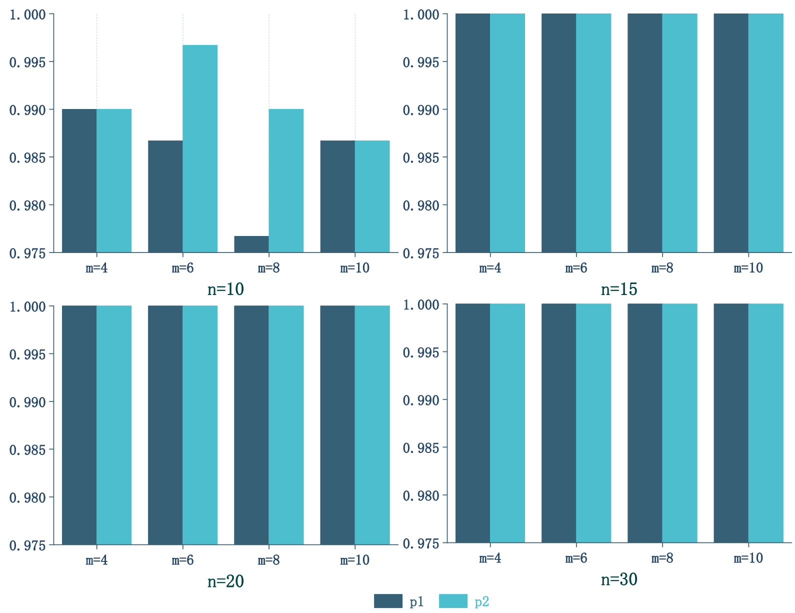

Figure 16: Probability bar charts corresponding to the fifth set of experiments

Table 17: Results of the sixth set of experiments ($x \sim | N\ [0,2.5] |$ and $\alpha \sim | N\ [1/2m, 2/m] |$)

Probability	n=10				n=15				n=20				n=30			
	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10
$\#\{P_1 \leq 0.05\}/K$	0.9967	0.9867	0.9800	0.9867	0.9967	1	1	0.9967	1	1	1	1	1	1	1	1
$\#\{P_2 \leq 0.05\}/K$	0.9967	0.9967	0.9900	0.9900	1	1	1	1	1	1	1	1	1	1	1	1
$\#\{P_2 < P_1\}/K$	0.2133	0.3567	0.3333	0.3967	0.1700	0.3167	0.3833	0.3100	0.0467	0.0700	0.0900	0.0900	0	0	0.0033	0
$\#\{P_2 = P_1\}/K$	0.6800	0.5433	0.5167	0.4600	0.7467	0.5967	0.5333	0.6133	0.9533	0.9200	0.9033	0.8900	1	1	0.9967	1
$\#\{P_2 > P_1\}/K$	0.1067	0.1000	0.1500	0.1433	0.0833	0.0867	0.0833	0.0867	0	0.0100	0.0067	0.0200	0	0	0	0

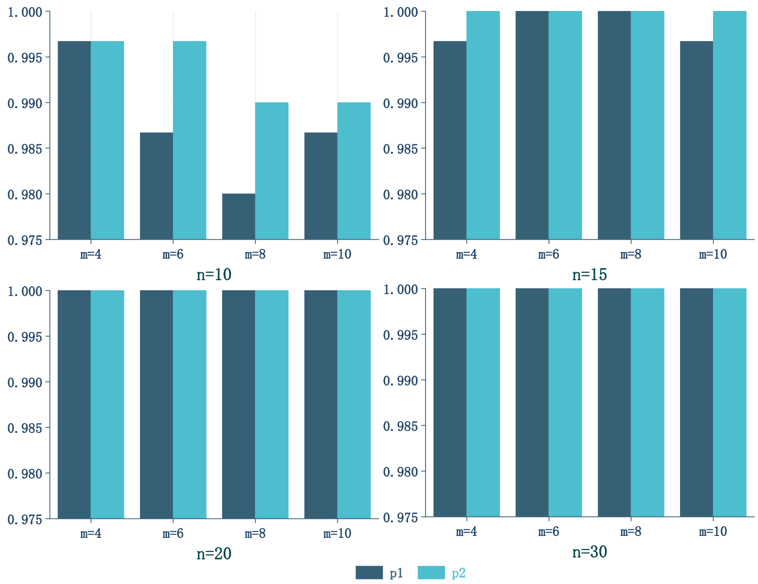

Figure 17: Probability bar charts corresponding to the sixth set of experiments

Table 18: Results of the seventh set of experiments ($x \sim U(1,6)$)

Probability	n=10				n=15				n=20				n=30			
	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10
$\#\{P_1 \leq 0.05\}/K$	1	1	1	0.9967	1	1	1	1	1	1	1	1	1	1	1	1
$\#\{P_2 \leq 0.05\}/K$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$\#\{P_2 < P_1\}/K$	0.0733	0.0767	0.1000	0.1000	0.0200	0.0400	0.0300	0.0233	0	0	0	0	0	0	0	0
$\#\{P_2 = P_1\}/K$	0.8933	0.8700	0.8500	0.8567	0.9767	0.9500	0.9567	0.9633	1	1	1	1	1	1	1	1
$\#\{P_2 > P_1\}/K$	0.0333	0.0533	0.0500	0.0433	0.0033	0.0100	0.0133	0.0133	0	0	0	0	0	0	0	0

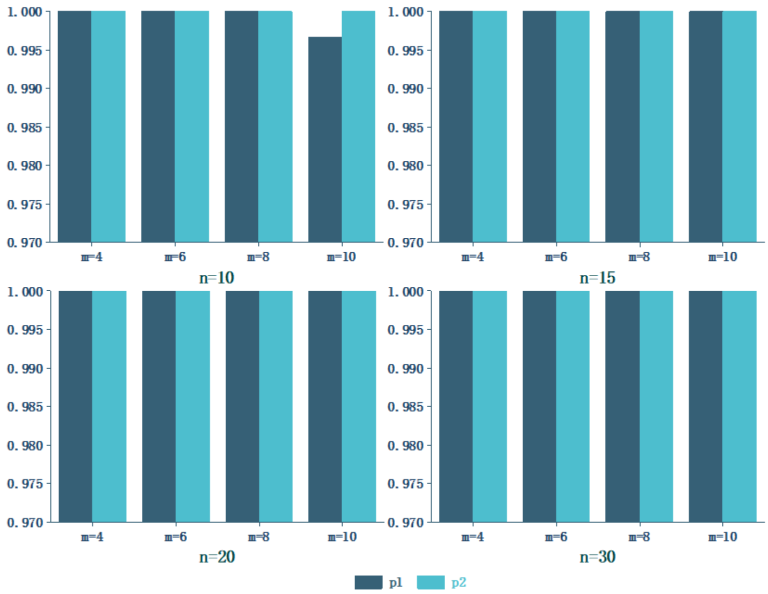

Figure 18: Probability bar charts corresponding to the seventh set of experiments

Table 19: Results of the eighth set of experiments ($x \sim U(1,6)$ and $\alpha \sim U(1/3m, 3/m)$)

Probability	n=10				n=15				n=20				n=30			
	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10
$\#\{P_1 \leq 0.05\}/K$	1	1	1	0.9967	1	1	1	1	1	1	1	1	1	1	1	1
$\#\{P_2 \leq 0.05\}/K$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$\#\{P_2 < P_1\}/K$	0.0733	0.0800	0.0767	0.0667	0.0233	0.0233	0.0133	0.0233	0	0.0033	0	0	0	0	0	0
$\#\{P_2 = P_1\}/K$	0.8867	0.8600	0.8700	0.8700	0.9633	0.9633	0.9800	0.9533	1	0.9967	1	1	1	1	1	1
$\#\{P_2 > P_1\}/K$	0.0400	0.0600	0.0533	0.0633	0.0133	0.0133	0.0067	0.0233	0	0	0	0	0	0	0	0

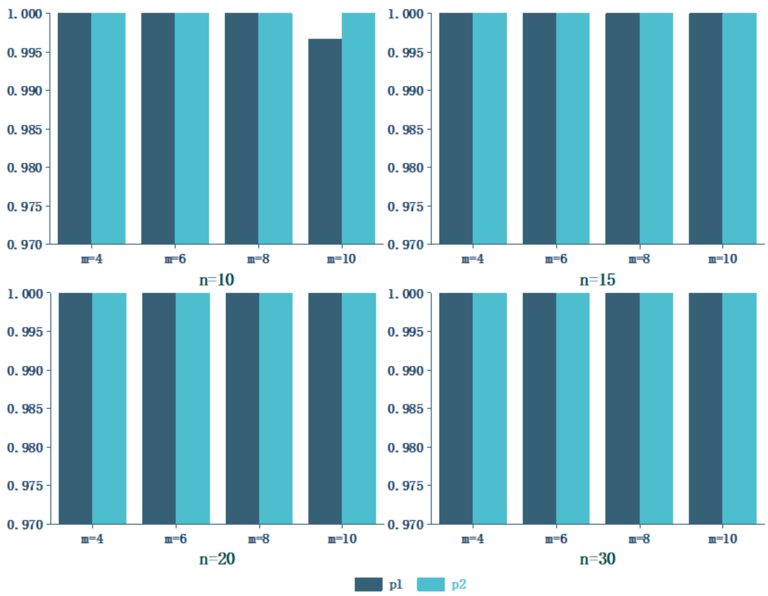

Figure 19: Probability bar corresponding to the eighth set of experiments

Appendix 3: Results of simulation where parameters are changed under IRS.

Table 20: Results of the ninth set of experiments (x~| N [0,2.5] |)

Probability	n=10				n=15				n=20				n=30			
	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10
$\#\{P_1 \leq 0.05\}/K$	1	1	1	1	1	1	1	0.9967	1	1	1	1	1	1	1	1
$\#\{P_2 \leq 0.05\}/K$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$\#\{P_2 < P_1\}/K$	0.2000	0.3200	0.3567	0.4233	0.1533	0.2200	0.3333	0.3767	0.0100	0.0367	0.0533	0.0600	0	0	0	0
$\#\{P_2 = P_1\}/K$	0.6533	0.5467	0.4933	0.4733	0.7867	0.7100	0.5700	0.5333	0.9800	0.9500	0.9267	0.9300	1	1	1	1
$\#\{P_2 > P_1\}/K$	0.1467	0.1333	0.1500	0.1033	0.0600	0.0700	0.0967	0.0900	0.0100	0.0133	0.0200	0.0100	0	0	0	0

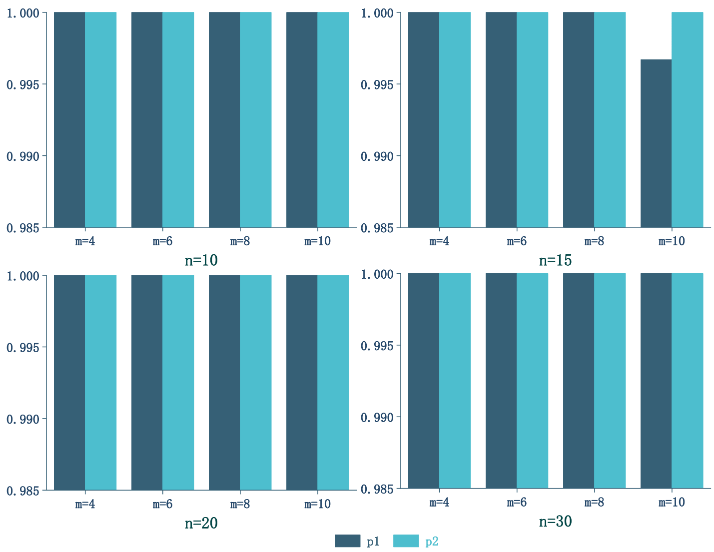

Figure 20: Probability bar charts corresponding to the ninth set of experiments

Table 21: Results of the tenth set of experiments ($x \sim |N[0,2.5]|$ and $\alpha \sim |N[1/2m, 2/m]|$)

Probability	n=10				n=15				n=20				n=30			
	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10
$\#\{P_1 \le 0.05\}/K$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$\#\{P_2 \le 0.05\}/K$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$\#\{P_2 < P_1\}/K$	0.1633	0.2867	0.3000	0.3900	0.1433	0.2300	0.2900	0.3400	0.0233	0.0767	0.0900	0.1033	0	0	0	0
$\#\{P_2 = P_1\}/K$	0.7133	0.6067	0.5733	0.4200	0.7800	0.6800	0.6300	0.5733	0.9700	0.9100	0.8933	0.8833	1	1	1	1
$\#\{P_2 > P_1\}/K$	0.1233	0.1067	0.1267	0.1900	0.0767	0.0900	0.0800	0.0867	0.0067	0.0133	0.0167	0.0133	0	0	0	0

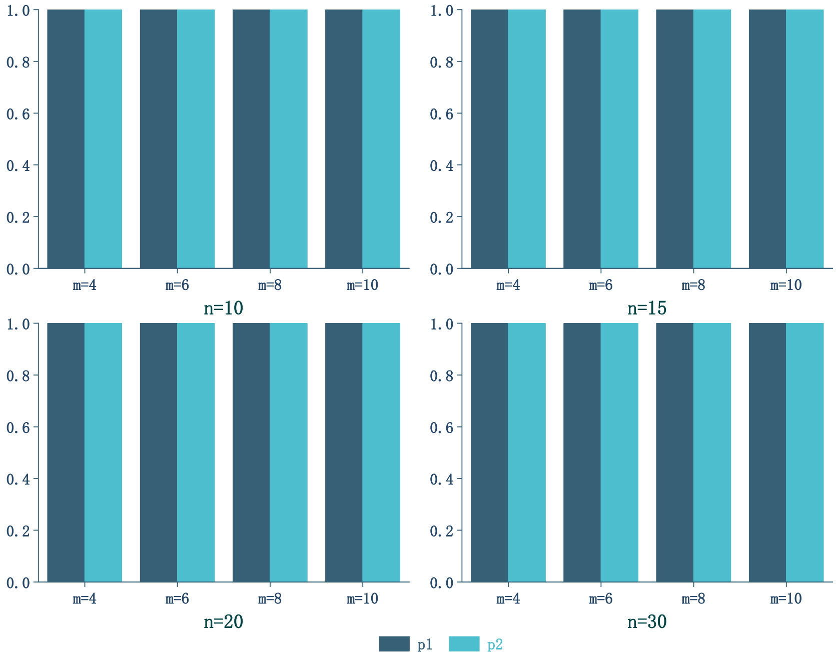

Figure 21: Probability bar charts corresponding to the tenth set of experiments

Table 22: Results of the eleventh set of experiments ($x \sim U(1,6)$)

Probability	n=10				n=15				n=20				n=30			
	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10
$\#\{P_1 \le 0.05\}/K$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$\#\{P_2 \le 0.05\}/K$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$\#\{P_2 < P_1\}/K$	0.1000	0.1000	0.1000	0.0767	0.0100	0.0233	0.0167	0.0100	0	0	0	0	0	0	0	0
$\#\{P_2 = P_1\}/K$	0.8500	0.8000	0.8400	0.8700	0.9667	0.9600	0.9767	0.9633	1	1	0.9967	1	1	1	1	1
$\#\{P_2 > P_1\}/K$	0.0500	0.1000	0.0600	0.0533	0.0233	0.0167	0.0067	0.0267	0	0	0.0033	0	0	0	0	0

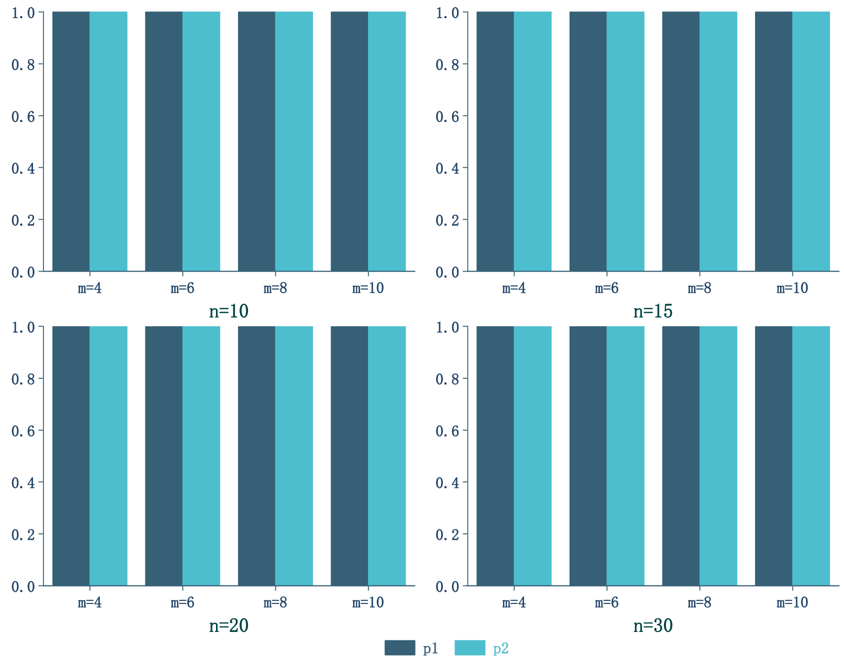

Figure 22: Probability bar charts corresponding to the eleventh set of experiments

Table 23: Results of the twelfth set of experiments ($x \sim U(1,6)$ and $\alpha \sim U(1/3m, 3/m)$)

Probability	n=10				n=15				n=20				n=30			
	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10	m=4	m=6	m=8	m=10
$\#\{P_1 \leq 0.05\}/K$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$\#\{P_2 \leq 0.05\}/K$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$\#\{P_2 < P_1\}/K$	0.0800	0.0300	0.0933	0.0800	0.0200	0.0267	0.0200	0.0333	0	0	0	0	0	0	0	0
$\#\{P_2 = P_1\}/K$	0.9000	0.9600	0.8800	0.8600	0.9667	0.9500	0.9700	0.9467	1	1	1	1	1	1	1	1
$\#\{P_2 > P_1\}/K$	0.0200	0.0100	0.0500	0.0600	0.0133	0.0033	0.0200	0.0200	0	0	0	0	0	0	0	0

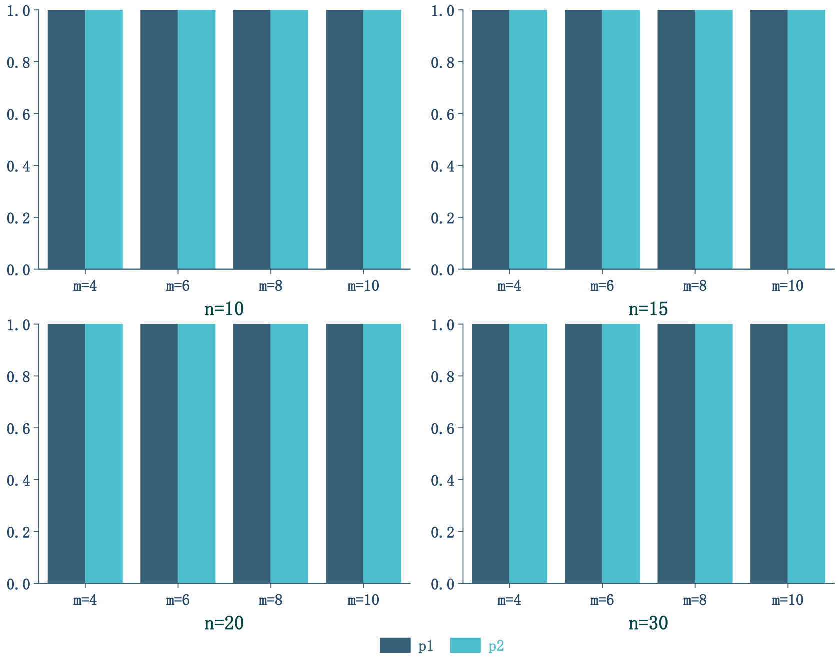

Figure 23: Probability bar corresponding to the twelfth set of experiments