Multi-Factor Policy Evaluation and Selection in the One-Sample Situation

Chien-Ming Chen

ERIM REPORT SERIES *RESEARCH IN MANAGEMENT*	
ERIM Report Series reference number	ERS-2008-084-LIS
Publication	December 2008
Number of pages	33
Persistent paper URL	http://hdl.handle.net/1765/14275
Email address corresponding author	cchen@rsm.nl
Address	Erasmus Research Institute of Management (ERIM) RSM Erasmus University / Erasmus School of Economics Erasmus Universiteit Rotterdam P.O.Box 1738 3000 DR Rotterdam, The Netherlands Phone: + 31 10 408 1182 Fax: + 31 10 408 9640 Email: info@erim.eur.nl Internet: www.erim.eur.nl

Bibliographic data and classifications of all the ERIM reports are also available on the ERIM website: www.erim.eur.nl

ERASMUS RESEARCH INSTITUTE OF MANAGEMENT

REPORT SERIES

RESEARCH IN MANAGEMENT

ABSTRACT AND KEYWORDS	
Abstract	Firms nowadays need to make decisions with fast information obsolesce. In this paper I deal with one class of decision problems in this situation, called the "one-sample" problems: we have finite options and one sample of the multiple criteria with which we use to evaluate those options. I develop evaluation procedures based on bootstrapping DEA (Data Envelopment Envelopment) and the related decision-making methods. This paper improves the bootstrap procedure proposed by Simar and Wilson (1998) and shows how to exploit information from bootstrap outputs for decision-making.
Free Keywords	multiple criteria, bootstrap, data envelopment analysis, parametric transformation, R&D project, supplier selection
Availability	
Classifications	

Multi-factor policy evaluation and selection in the one-sample situation

Chien-Ming Chen
cchen@rsm.nl
Rotterdam School of Management,
Erasmus University, the Netherlands

Draft version: Nov 2008.

Abstract

Firms nowadays need to make decisions with fast information obsolesce. In this paper I deal with one class of decision problems in this situation, called the "one-sample" problems: we have finite options and one sample of the multiple criteria with which we use to evaluate those options. I develop evaluation procedures based on bootstrapping DEA (Data Envelopment Envelopment) and the related decision-making methods. This paper improves the bootstrap procedure proposed by Simar and Wilson (1998) and shows how to exploit information from bootstrap outputs for decision-making.
Keywords: Multiple criteria; Bootstrap; Data envelopment analysis; Parametric transformation; R&D project & supplier selection.

1 Introduction

Firms in today's supply chains are faced with increasing pressure to respond to market changes and potential problems in realtime. This competition also not only induces frequent product introductions and technology advance and increased customer welfare (Petrin, 2002), but also shortened product live cycles. Constant product innovation and rapid introductions necessitate fast decision-making and planning under a high level of uncertainty; Bourgeois and Eisenhardt (1988) point out that in the high-velocity environment (for example in the computer industry), the incessantly changing market can make information inaccurate, unavailable or obsolete.

One way to deal with uncertainty in decision-making is using Ranking & Selection (R&S) methods, in which we used known stochastic processes to represent the uncertain factors in the environment (i.e., *parametric* problems), and hence we can approximate the statistics of interest with the help of random number generators. In high-velocity, complex, or unstable environments, however, it is not always straightforward to prescribe

probability distributions for factors beyond our control. The need for fast decisions usually disallow us the luxury to conduct a large-scale and comprehensive survey or experiment, and therefore the attempt to rely on comprehensive data become impractical. On the other hand, the decision based on the robust optimization school of thought can be too conservative (see, e.g., Ben-Tal and Nemirovski, 1999, and Bertsimas and Thiele, 2004).

In this paper, I deal with one class of decision problems in the "high-velocity" environment: in these problems, we have only one observation (or estimation) of the inputs and outputs of decision alternatives, and we want to distribute limited resources among available options. The decision-making process is considered a two-step "evaluate-then-decide" process. For instance, we rank different designs according to certain criteria first, and then select best or the best few designs. A firm may have many parallel projects competing in the new product development, but only the best or the best few only will be selected—because of the consideration of product variety, market saturation, or simply the budget constraint. So the resource constraint may be either expressed as the total amount of resources available, or as the maximum number of alternatives selected.

To evaluate different options, I develop bootstrap algorithms based on Data Envelopment Analysis (DEA) to tackle the above problem. DEA is a nonparametric approach to measure relative efficiency of systems that use multiple inputs to produce multiple outputs. DEA is also has its root in production economics, which provide a firm base necessary for the development of bootstrap algorithms (Coelli, 2005). Simar and Wilson (1998) first develop the bootstrap algorithm based on the BCC DEA model (Banker et al., 1984). Based on one sample of input and output variables, our method can be used to approximate the true efficiency distributions of all evaluated units. From the empirical result that will be shown later, we will see that the DEA model, as an efficiency estimator, tend to be volatile for certain efficient units. Hence these efficient units would be considered "risky" in their performance, which gives a contrary picture to the underlying performance. I rectify this problem by modifying the cross-efficiency method Sexton et al. (1986) to develop a new efficiency estimator.

In addition, Simar and Wilson (1998) do not indicate the decision-making procedure based on the bootstrap outputs. This is a critical gap because the goal of evaluation is to facilitate subsequent decision-making and planning activities. Therefore I propose two approaches to extract information from the evaluation outcomes (i.e., the bootstrap distributions). First I use the bootstrap distributions to estimate the mean-variance statistics and use the estimates to solve the mean-variance formulation of the project selection problem. We can then achieve the optimal mean-variance tradeoff in the efficiency of project investment portfolios. Next I develop a normalization algorithm to transform the nonparametric bootstrap distributions into normal ones. The normalization procedure allows us to apply a wide range of statistical methods that require a normality

assumption.

In the next section, I will introduce relevant backgrounds about the DEA model and the bootstrap method. In Sec. 3 and 4, I elaborate on the probability models and bootstrap algorithms constructed based on the original and modified cross-efficiency method, respectively. In Sec. 5, the distinction between these two models is compared through an illustration based on empirical R&D project data. In Sec. 6, we apply the mean-variance formulation to the bootstrap distributions to solve project selection and budgeting from an efficiency viewpoint. In Sec. 7, I use batch means and the method of multiple statistical testings to develop a normalization algorithm for the bootstrap distributions. The final section provides a summary of this paper.

2 Models and preliminaries

This section introduces the Cross-Efficiency (CE) method, and review the statistical aspects of efficiency measuring in this section. The bootstrap principle is also briefly discussed.

2.1 DEA model and CE method

Almost all production processes involve multiple production factors. A customary approach to deal with multiple factors in evaluation is to assign weights to each factor to form a single performance indicator. In many situations, however, it can be difficult to find weights that can properly represent the relative importance of all factors, and thus we are not capable of separating the real inefficiency from the effect due to weight specification in evaluation result (Cooper et al., 2006). DEA has been widely used to evaluate the relative performance of units with multiple inputs and outputs. As opposed to assigning fixed weights to inputs and outputs, DEA allows each evaluated unit to determine weights to optimize the evaluated unit's efficiency score.

Suppose there are n decision-making units (DMUs) under evaluation. For an arbitrary DMU k, it requires inputs $X_k = [X_{k1}, \ldots, X_{ki}]$ to yield outputs $Y_k = [Y_{k1}, \ldots, Y_{kj}]$. The input-oriented CCR efficiency is defined to be the optimal value of the fractional linear problem (Charnes et al., 1978):

$$\max \sum_{q=1}^{j} \mu_{1q} Y_{1q} \Big/ \sum_{p=1}^{i} \nu_{1p} X_{1p} \tag{1a}$$

$$\text{subject to} \quad \sum_{q=1}^{j} \mu_{1q} Y_{kq} - \sum_{p=1}^{i} \nu_{1p} X_{kp} \leq 0, \ k = 1, \ldots, n, \tag{1b}$$

$$\mu_{1q}, \nu_{1p} \geq 0, \ p = 1, \ldots, i, \text{ and } q = 1, \ldots, j. \tag{1c}$$

The objective function of (1) is the weighted ratio of output and input factors, which conforms to the classical definition of productivity. Therefore (1) will maximize the evaluated DMU's efficiency by choosing some nonnegative weights ν_{1p}'s and μ_{1q}'s. The value of ν_{1p} and μ_{1q} can be interpreted as the relative importance of the variables in the evaluation process. Constraint (1b) ensures that the weights will not render the efficiency scores of any DMU larger than one.

By normalizing the the denominator of (1a), we obtain an equivalent LP to problem (1):

$$\max \theta_1 = \sum_{q=1}^{j} \mu_{1q} Y_{1q} \tag{2a}$$

$$\text{subject to} \quad \sum_{p=1}^{i} \nu_{1p} X_{1p} = 1, \tag{2b}$$

$$\sum_{q=1}^{j} \mu_{1q} Y_{kq} - \sum_{p=1}^{i} \nu_{1p} X_{kp} \leq 0, \ k = 1, \ldots, n, \tag{2c}$$

$$\mu_{1q}, \nu_{1p} \geq 0, \ p = 1, \ldots, i, \text{ and } q = 1, \ldots, j. \tag{2d}$$

The evaluation process completes after we repeatedly solve problem (2) for n times, each time for one DMU; i.e., we substitute X_{kp} and Y_{kq} in (2a) and (2b) in each round. DMUs that obtain an efficient score of one are called *efficient*, and *inefficient* otherwise.

Now we are ready to define the cross-efficiency. Denote the optimal solution to (2) for DMU k by the pair $(\hat{\boldsymbol{\nu}}_k, \hat{\boldsymbol{\mu}}_k)$, where $\hat{\boldsymbol{\nu}}_k = [\hat{\nu}_{k1}, \ldots, \hat{\nu}_{ki}]$ and $\hat{\boldsymbol{\mu}}_k = [\hat{\mu}_{k1}, \ldots, \hat{\mu}_{kj}]$. So after the DEA evaluation, we can obtain n sets of weights corresponding to n DMUs. The *cross-efficiency* (CE) for DMU k is then defined to be

$$\text{CE}_k = n^{-1} \sum_{r=1}^{n} \frac{\sum_{q=1}^{j} \hat{\mu}_{rq} Y_{kq}}{\sum_{p=1}^{i} \hat{\nu}_{rp} X_{kp}}. \tag{3}$$

In other words, the CE is the average of efficiency scores associated with optimal weights determined by all DMUs. (3) reveals two main features of CE. In DEA models, it often occurs that multiple DMUs are identified as efficient. It is also well-known that the discrimination power (i.e., the proportion of DMUs being identified as efficient) will be even weaker when we only have relatively small sample, as compared with the number of input and output variables[1]. It is straightforward to show that $\text{CE}_k \leq \theta_k$, and we can almost always use CE scores to obtain non-tied rankings of DMUs because the CE score takes in weights given by all evaluated DMUs. Thus we can use the CE method

[1]This is also known as *the curse of dimension* (see also the discussion in the next subsection). A rule of thumb for practitioners is that the number of projects should be at least more than the square of the number of input and output variables combined; see, e.g., Dyson et al. (2001).

to effectively increase the discrimination in the evaluation result (Adler et al., 2002). In addition, the CE method can also mitigate the influence of flexible weight selection obtained from (2). Finally, Doyle and Green (1994) argue that the CE method is a democratic evaluation process, as compare to traditional DEA models, because DEA models determine the efficiency score according to the perspective of the evaluated DMU only. Therefore, CE has been extensively applied in many evaluation problems; see, among others, Oral et al. (1991); Doyle and Green (1994); Shang and Sueyoshi (1995); Green et al. (1996); Chen (2002); Talluri and Narasimhan (2004); Liang et al. (2008b).

Determination of unique optimal weights

One implementation issue of the CE method is that the optimal solution to (2) is likely to be degenerate and thus non-unique. The solution will then depend on the optimization packages used (Despotis, 2002). Several studies provide antidotes to the problem. Doyle and Green (1994) propose an auxiliary procedure to determine a set of unique weights, given the evaluated DMU's DEA efficiency score. Specifically, Doyle and Green use the following optimization problem:

$$\min \sum_{q=1}^{j}\sum_{k=2}^{n} \mu_{1q}Y_{kq} \Big/ \sum_{p=1}^{i}\sum_{k=2}^{n} \nu_{1q}X_{kq} \tag{4a}$$

$$\text{subject to} \quad \sum_{q=1}^{j} \mu_{1q}Y_{1q} \Big/ \sum_{p=1}^{i} \nu_{1q}X_{1q} - \theta_1 = 0, \tag{4b}$$

$$\sum_{q=1}^{j} \mu_{1q}Y_{kq} - \sum_{p=1}^{i} \nu_{1p}X_{kp} \leq 0,\ k = 1, \ldots, n, \tag{4c}$$

$$\mu_{1q}, \nu_{1p} \geq 0,\ p = 1, \ldots, i,\ q = 1, \ldots, j. \tag{4d}$$

In problem (4), the objective function is formulated as the weighted output-input ratio of the other DMUs except DMU 1 (the evaluated DMU), while (4a) ensures the weights are optimal to problem (2). So problem (4) searches, among the optimal solutions to (2), the set of weights that minimizes the weighted output-input ratio related to other DMUs. More specifically, constraint (4b) makes sure that the efficiency score θ_1 obtained from (2) is maintained. Problem (4) is thus called the *aggressive* formulation due to its objective function. Sexton et al. (1986) develop a similar method, called the *benevolent* formulation, that instead maximizes the same objective as in (4a). See also Liang et al. (2008a) and Liang et al. (2008b) for variant methods to derive the weights.

2.2 A statistical view of DEA models

Efficiency measuring has been deeply rooted in production economics. Over the years, various econometrics methods, parametric or nonparametric, have been developed to "estimate" efficiency in different contexts and under different assumptions. Charnes et al. (1978) coined the DEA model in the form of an LP formulation, which can also be regarded as a nonparametric efficiency estimator. However, many researchers have since regarded DEA as a deterministic approach, which seems to detach DEA from its statistical implications; see also the discussion in Simar and Wilson, 1999. From a statistical viewpoint, the efficiency estimates from DEA hinges closely on its nonparametric estimate of the unknown production function, and the production frontier estimate of DEA is susceptible to finite sample errors and sample variations (Kneip et al., 1998). Therefore an understanding of the statistical aspects of the scores can shed further light on the precision and confidence of the evaluation results. We begin this section by reviewing the statistical aspects of DEA and the bootstrap principle.

Properties of DEA as an efficiency estimator

The development of DEA as an efficiency estimator stems from the framework of conventional production economics (Coelli, 2005). In the framework, we draw samples containing the productive information of firms and use it to estimate the relation, which we call a *production function*, between inputs and outputs in production. Simultaneously, we can deduce the technical inefficiency of firms by comparing the samples with the production function. Following this construction, DEA provides a nonparametric and piecewise linear estimation of the production function. The exploration of statistical properties of DEA commences with the pioneer work of Banker (1993). He proves the consistency of DEA in the univariate case (single input with multiple outputs) for any concave and monotone production functions. He also proposes an asymptotic test procedure for statistical inferences on efficiencies estimated by DEA models (see also Banker and Chang (1995) for a comparison between different testing methods). Korostelev et al. (1995) and Kneip et al. (1998) give the convergence rate of DEA estimators in more general settings, and show that, under a fixed sample size, the convergence rate decreases exponentially as the number of input/output variables increase.

Simar and Wilson (1998) use a nonparametric smoothed bootstrap method to approximate the sampling distribution of DEA efficiency estimates. However, they did not give the standard error and confidence intervals of the efficiency estimate in their paper. Simar and Wilson (1999) apply a similar bootstrap method to the Malmquist index. Further, Simar and Wilson (2000a) extend the method proposed in Simar and Wilson (1998) to a more general case which allows the efficiency distributions to be heterogeneous. Simar and Wilson (2000b) provide a detailed overview about recent developments for the sta-

tistical analysis of DEA. It should be stressed that the statical developments so far are exclusively concerned about the "radial" or "envelopment" DEA model, i.e., the primal formulation to problem (2). The view developed later in this paper is however related to the multiplier model (2), which is still an unexplored area in the literature.

2.3 Bootstrap preliminaries

The primary goal in statistical inference is to use a random sample to infer about parameters associated with an unknown population F_0. To draw a statistical conclusion based on a sample, it requires sufficient information about the sampling distribution of an appropriate statistic (or an *estimator*). Accessibility of sampling distributions, however, relies heavily on both the analytical properties of F_0 and the mathematical structure of the estimator. Therefore, only under special assumptions on F_0 can we give the analytical description about the sampling distributions of certain estimators. This would impose great restrictions on the applicability of statistical analysis in many real-world problems.

Bootstrapping is a collection of computational methods that can overcome the above issues. Under fairly general conditions, we can use bootstrap methods to approximate sampling distributions by resamplings from the obtained sample. Efron and Tibshirani (1993) is the classic introduction to bootstrap methods, while Hall (1992) gives a comprehensive theoretical treatment of bootstrap theories. Below I briefly introduce the bootstrap principle for nonparametric problems.

Bootstrap methods is mainly used to approximate sampling distributions that are impossible for analytical representation. The basic principle is to use the observed sample as an estimate of the population of interest[2]. We can then draw bootstrap samples repetitively from the estimated population to approximate the sampling distribution of interest.

The way in which we construct the estimated population will determine the bootstrap to be parametric or nonparametric. In parametric bootstrap methods, the obtained sample will used to estimate the parameters associated with the predetermined probability distribution; bootstrap samples are then drawn from the distribution given the estimated parameters. Nonparametric bootstrap methods, on the other hand, do not require a distributional assumption, and will be introduced next.

Let $(x_1, x_2, \ldots, x_n)$ be a random sample drawn from the unknown F_0, namely

$$(x_1, x_2, \ldots, x_n) \overset{\text{iid}}{\sim} F_0. \tag{5}$$

In the nonparametric method, F_1 represents the empirical distribution of the sample

[2]Efron and Tibshirani (1993) contrast the true population and the estimated population by referring them as the "real world" and "bootstrap world," respectively; Hall (1992) vividly exemplifies the two populations by Russian matryoshka dolls.

of size n drawn from F_0. Specifically, F_1 is constructed by allocating $1/n$ probability mass to each x_i in the random sample:

$$F_1 := P(x = x_i) = \frac{\#(x_j = x_i, \text{ for } j = 1, \ldots, n)}{n}, \text{ for } i = 1 \text{ to } n, \tag{6}$$

#() being the cardinality function.

Let $(x_1^*, x_2^*, \ldots, x_n^*)$ represent the random sample from F_1:

$$(x_1^*, x_2^*, \ldots, x_n^*) \overset{\text{iid}}{\sim} F_1. \tag{7}$$

By sampling from F_1, we can obtain the bootstrap distribution of the estimator of interest g[3]. The relation resultant distribution F_2 represents an approximation to the sampling distribution given F_0.

The bootstrap principle refers to the assumption that the relationship between F_1 and F_2 is a close resemblance to the relation between F_0 and F_1. Since we have full knowledge of the empirical distribution (6), we can use Monte Carlo simulation to approximate F_2 based on F_1.

3 Bootstrapping the CE model

Simar and Wilson (1998) pioneered the use of a smoothed bootstrap procedure to estimate the sampling distribution of inefficiency score from the DEA BCC radial model (Banker et al., 1984). They also detail the data generating process from the perspective of a DEA radial model. Simar and Wilson (1999) extend the idea to bootstrapping the Malmquist productivity index.

In this section, we introduce the bootstrap algorithm for the CE model. The bootstrap algorithm helps us obtain approximated distributions of CE scores, which can be used for inferences and comparisons on the CE scores of sampled DMUs. The methodology developed in this paper enables one to assess the peer-appraisal efficiency and the underlying efficiency variations. More specifically, I use a nonparametric bootstrap method to approximate the sampling distribution of CE scores.

The bootstrap framework developed in this paper has several distinctive features, as compared with Simar and Wilson (1998). First, the CE model promotes an encompassing assessment philosophy, as described earlier. Therefore the efficiency structure is different from that considered in the literature. Mathematically, the proposed bootstrap method views the problem from the dual programming perspective of the radial DEA model (i.e., the DEA multiplier model) as in Simar and Wilson (1998). So our model corresponds to a "multi-parameter" estimation problem; see, e.g., Efron (1987).

[3]See Efron and Tibshirani (1993) for suggestions about the appropriate values of B for different purposes of bootstrapping.

3.1 Probability model and bootstrap algorithm

For all DMUs, the feasible region of problem (1) is a polyhedron associated with the sample size n—that also means that the DEA results will depend on our estimate of the polyhedron. For now, we assume that the sequence of polyhedra converges downwards to a nonempty region of permissible weights, as the sample size tends to infinity:

$$(\boldsymbol{\nu}, \boldsymbol{\mu}) := \bigcap_{k=1}^{\infty} \Big\{ (\nu_{kp}, \mu_{kq}) : \sum_{q=1}^{j} \mu_{kq} Y_{kq} - \sum_{p=1}^{i} \nu_{kp} X_{kp} \leq 0 \Big\} \subset \Re_{+}^{i \times j}. \tag{8}$$

By (1b), the estimate of (8) based on a sample of size n can be represented by

$$(\hat{\boldsymbol{\nu}}_{\boldsymbol{n}}, \hat{\boldsymbol{\mu}}_n) = \bigcap_{k=1}^{n} \Big\{ (\nu_{kp}, \mu_{kq}) : \sum_{q=1}^{j} \mu_{kq} Y_{kq} - \sum_{p=1}^{i} \nu_{kp} X_{kp} \leq 0 \Big\}. \tag{9}$$

We then arrive at the following convergence result:

Theorem 1. *The set sequence defined in* (9) *is non-increasing and will converge in probability to* (8)*; i.e.,*

$$(\hat{\boldsymbol{\nu}}_{\boldsymbol{n}}, \hat{\boldsymbol{\mu}}_n) \xrightarrow{P} (\boldsymbol{\nu}, \boldsymbol{\mu}).$$

The above proposition can be deduced from the the primal, radial DEA formulation by LP dualities; see Banker (1993), Korostelev et al. (1995), Kneip et al. (1998) and Simar and Wilson (2000b) for the primal construction of the efficiency estimation problem[4]. Therefore, we can regard (9) as a finite sample approximation of the *true* polyhedron $(\boldsymbol{\nu}, \boldsymbol{\mu})$ in (8). Recall that after solving (2) for each DMU, we obtain a set of optimal input-output weight vectors $(\hat{\mu}_{kq}, \hat{\nu}_{kp})$, $k = 1, \ldots, n$, by which we can calculate the CE scores. Following the notations introduced earlier, we construct the empirical distribution of DMUs (F_1) by putting probability mass $1/n$ on each DMU[5]. Then we can generate bootstrap samples by drawing samples from F_1:

$$F_1 \rightarrow \big((X_1, Y_1)^*, \ldots, (X_n, Y_n)^* \big), \tag{10}$$

from which we can obtain bootstrap replications by computing $\widehat{CE}_k^*$ as in (3), for $k = 1, \ldots, n$. By repeatedly sampling from F_1, we can approximate the distributions of CE scores. The process of deriving $\widehat{CE}_k^*$ is referred to as the multiparameter problem in the literature; see, e.g., Efron and Tibshirani (1986) and Efron (1987) for further discussions. Note that the equivalence of problem (1) and problem (2) implies that weight estimates obtained from (2) must be optimal for (1) as well.

[4]There are few assumptions necessary for the convergence results in the primal efficiency framework: (1) i.i.d. sampling, (2) a convex production set, (3) efficient units will be observed with probability one as $n \to \infty$.

[5]This approach is analogous to bootstrapping pairs in the regression analysis; see, e.g., Chap. 9 of Efron and Tibshirani (1993).

We summarize the above bootstrap procedure in Algorithm 1:

Algorithm 1 CE bootstrap algorithm

1: **for** $b = 1$ to B **do**
2: sample with replacement from the empirical input-output pairs to generate bootstrap samples $(X_{kb}, Y_{kb})^*$, $k = 1$ to n.
3: **if** the bootstrap sample is degenerate (all input-output pairs in the resample are identical) **then**
4: set $\hat{w}^*_{kb} = (\hat{\nu}^*_{kb}, \hat{\mu}^*_{kb})$, where $\hat{\nu}^*_{kbp} = 1 \Big/ \sum_{p=1}^{i} X_{1p}$, $\hat{\mu}^*_{kbq} = 1 \Big/ \sum_{q=1}^{j} Y_{1q}$, $\forall k, p, q$.
5: **else**
6: estimate the input-output weight vector $\hat{w}^*_{kb}$ by using models (2) and (4), for $k = 1$ to n.
7: compute $\widehat{CE}^*_{kb}$ according to (3), for $k = 1$ to n.
8: **end if**
9: **end for**
10: obtain the bootstrap distribution of $\widehat{CE}^*_k$, for $k = 1$ to n.

In STEP 2, we resample the observed input/output firms to estimate the weight vector in STEP 6. STEPs 3 and 4 are developed to avoid the degenerate case, where the resample consist of only replicas of one specific empirical sample and thus we would be unable to determine the optimal weight vector uniquely. Hence in the algorithm we regulate that, if the bootstrap sample is degenerate, then the weights are specified as shown in STEP 4, in which input and output variables are assigned equal weights. The probability of obtaining a degenerate resample, however, is fairly insignificant in practice. On the other hand, if the resample is non-degenerate, we can proceed to compute the bootstrap weights by invoking models (2) and (4) for the resample. Provided the bootstrap weights, the bootstrap distributions of CE are straightforward (STEPs 7 and 10).

Finally, the model of efficiency estimation introduced in this section is based on the DEA multiplier model (1). We can make an interesting comparison between our and the estimation based on the DEA radial model. In the latter model, DEA provides a downward biased estimate of the true production frontier under the concavity and monotone assumptions; i.e., the estimated estimated production feasible set is a subset of the true one. Therefore, the DEA model tends to give an overoptimistic estimate about efficiencies. In our model, the estimated feasible region of weights (9) is an overestimate of (8). Hence the actual freedom of weight selection should be smaller, which means that the efficiency will tend to be overestimated. More specifically, model (1) tends to get *tighter* bounds on weights when the sample size increases—for (2c), once a larger sample is involved, more constraints will then be introduced in the problem. This can be contrasted with the radial DEA model, in which the production feasibility set cannot be

reduced by the addition of new observations. The above relation interestingly corresponds to the primal-dual relationship between the two problem formulations.

4 The revised CE method

We introduced the bootstrap algorithm regarding the CE method in the previous section. The CE method operates in parallel with the DEA model, which is sensitive to variations of the extremely efficient DMUs (see, e.g., Zhu (2001)). In the numerical results shown later, we will see evidence that the bootstrap distributions of efficient DMUs have exceptionally wide spreads. In this section, we propose an alternative CE model that can circumvent this problem but still retain the merit of the CE method: Better discrimination and a democratic evaluation scheme. Based on the revised CE method (RCE), I formulate a bootstrap algorithm that can offer different insights on efficiency variations.

4.1 Revised Cross-Efficiency (RCE) method: definition

In the CE method introduced in previous sections, the efficiency score is computed as an average of the self- and peer-evaluations scores, which make use of the self- and peer-evaluations weights, respectively. Specifically, the RCE score is defined by

$$RCE_k(F_0) = \sum_{q=1}^{j} \mathbb{E}\left\{\mu_q\right\} Y_{kq} \Big/ \sum_{p=1}^{i} \mathbb{E}\left\{\nu_p\right\} X_{kp}, \tag{11}$$

where μ_q, ν_p are the weight distributions with respect to input p and output q; X_{kp}, Y_{kq} are known input and output levels, respectively. It is easy to show that, just like the CE score, it holds that the RCE score is bounded between 0 and 1, and it is less than or equal to the DEA score θ_k. In (11), the efficiency score is determined as the ratio of the virtual input to virtual output (1a), which is similar to (3). Unlike the CE method, however, the RCE method is defined on mean weights.

If we assume that input weight distributions are independent of output weight distributions, an unbiased estimator of the RCE based on a size n sample is

$$RCE_k(F_0) = \sum_{q=1}^{j} \hat{\bar{\mu}}_q^* Y_{kq} \Big/ \sum_{p=1}^{i} \hat{\bar{\nu}}_p^* X_{kp}, \tag{12}$$

where $\hat{\bar{\mu}}_q^*$ and $\hat{\bar{\nu}}_p^*$ are the estimates of $\bar{\mu}_q^*$ and $\bar{\nu}_p^*$, respectively.

Let us now apply the RCE method for a better understanding of its relation with the CE method. In the illustrative example, we evaluate 18 suppliers for outsourcing activities based on four criteria (two inputs and two outputs); the data set can be found in Kleinsorge et al. (1992). Table 1 lists the evaluation results of DEA, CE and RCE methods. The results show that, based on a similar spirit to the CE method, the RCE

method can also enhance discrimination power and provide scores and ranks close to those of the CE method (p=0.617 in the Wilcoxon sign-rank test).

Table 1: Suppliers' evaluation results for outsourcing (ranks in the parentheses)

Supplier	DEA CCR model	CE method	RCE method
1	0.9969 (9)	0.8687 (12)	0.8727 (11)
2	1.0000 (1)	0.9000 (10)	0.9022 (9)
3	1.0000 (1)	0.9591 (1)	0.9583 (1)
4	1.0000 (1)	0.9157 (8)	0.9185 (7)
5	0.9925 (10)	0.9040 (9)	0.9074 (8)
6	1.0000 (1)	0.9543 (3)	0.9558 (2)
7	0.9659 (15)	0.7993 (16)	0.8045 (15)
8	0.9799 (12)	0.8437 (15)	0.8379 (14)
9	0.9809 (11)	0.8557 (14)	0.8518 (13)
10	0.8593 (18)	0.7214 (18)	0.7152 (18)
11	0.8615 (17)	0.7491 (17)	0.7458 (17)
12	0.9255 (16)	0.8591 (13)	0.8597 (12)
13	1.0000 (1)	0.9573 (2)	0.9547 (3)
14	1.0000 (1)	0.9235 (7)	0.9218 (6)
15	1.0000 (1)	0.9518 (6)	0.9509 (5)
16	0.9787 (14)	0.8290 (16)	0.8117 (16)
17	0.9791 (13)	0.8966 (11)	0.8985 (10)
18	1.0000 (1)	0.9522 (4)	0.9526 (4)

4.2 Probability model and bootstrap algorithm

The probability model in RCE is different from what we considered in the previous section. Consider n DMUs under evaluation. For the evaluated DMU, $n-1$ other DMUs are randomly drawn from an unknown population F_0. Then for each evaluated DMU, the DEA model produces n weight vectors $(\hat{\boldsymbol{\omega}}_1, \ldots, \hat{\boldsymbol{\omega}}_n)$, where $\hat{\boldsymbol{\omega}}_k = (\hat{\nu}_{k1}, \ldots, \hat{\nu}_{ki}, \hat{\mu}_{k1}, \ldots, \hat{\mu}_{kj})$, for $k = 1, \ldots, n$. A straightforward estimator for the mean weight vector is the sample mean:

$$\bar{\boldsymbol{\omega}} = \sum_{k=1}^{n} \hat{\boldsymbol{\omega}}_k / n = (\bar{\nu}_{k1}, \ldots, \bar{\nu}_{ki}, \bar{\mu}_{k1}, \ldots, \bar{\mu}_{kj}),$$
$$\text{where } \bar{\nu}_q = \sum_{k=1}^{n} \hat{\nu}_q / n, \ \bar{\mu}_q = \sum_{k=1}^{n} \hat{\mu}_q / n. \tag{13}$$

Eqn. (13) leads to a plug-in estimator of RCE efficiency:

$$\widehat{RCE}_k = \frac{\sum_{q=1}^{j} \bar{\mu}_q Y_{kq}}{\sum_{p=1}^{i} \bar{\nu}_p X_{kp}} = \frac{\sum_{q=1}^{j} \sum_{k=1}^{n} \hat{\mu}_{kq} Y_{kq}}{\sum_{p=1}^{i} \sum_{k=1}^{n} \hat{\nu}_{kp} X_{kp}}. \tag{14}$$

Denote the probability distribution of $\boldsymbol{\omega}_k$ by F_0. Since F_0 is unknown, we can estimate

F_0 by its empirical distribution F_1, which assigns probability mass $1/n$ on $\hat{\boldsymbol{\omega}}_k$ for $k = 1$ to n.

Given F_1, we can generate bootstrap samples of weight vectors:

$$F_1 \rightarrow (\hat{\boldsymbol{\omega}}_1^*, \hat{\boldsymbol{\omega}}_2^*, \ldots, \hat{\boldsymbol{\omega}}_n^*). \tag{15}$$

Based on the bootstrap samples we can calculate $\bar{\boldsymbol{\omega}}^*$ and $\widehat{RCE}_k^*$ according to (13) and (15). By repeating this procedure, we obtain the bootstrap distribution of $\widehat{RCE}_k^*$, for $k = 1$ to n. We summarize the above procedure in Algorithm 2.

Algorithm 2 CE weight bootstrapping algorithm

1: estimate the weight vectors $(\hat{\boldsymbol{\omega}}_1, \hat{\boldsymbol{\omega}}_2, \ldots, \hat{\boldsymbol{\omega}}_n)$ by using model (2) and (4).
2: **for** $b = 1$ to B **do**
3: **for** DMU $k = 1$ to n **do**
4: sample with replacement from the empirical weight vectors $\hat{\boldsymbol{\omega}}_{b1}^*, \hat{\boldsymbol{\omega}}_{b2}^*, \ldots, \hat{\boldsymbol{\omega}}_{bn}^*$.
5: obtain a bootstrap sample $\bar{\boldsymbol{\omega}}_b^* = n^{-1} \sum_{k=1}^n \boldsymbol{\omega}_{bd}^*$ for $d = 1, \ldots, n$.
6: compute $\widehat{RCE}_{kb}^*$ according to Eqn. (14).
7: **end for**
8: **end for**

Finally, it is worth mentioning that the RCE bootstrap method is computationally more efficient than Algorithm 1 and the bootstrap method by Simar and Wilson (1998), although they are constructed based on different efficiency models. This is because computing RCE only requires few algebraic operations as shown in (14).

So far we have developed two bootstrap algorithms based on different definitions of cross-efficiencies. In the next section, we will apply the algorithms to empirical project performance data to see how the algorithms work out.

5 Empirical comparisons and results

Next we apply two bootstrap algorithms developed so far to empirical data, which consist of 37 R&D project proposals in the iron and steel industry of Turkey. The data first appeared in Oral et al. 1991 and were also used in Liang et al. (2008b) (see Tables 5). Each project requires a certain amount of budget (second column) and has five contribution indicators (column 2 to 7). Our objective is to evaluate project performance based on the required resources and intended contribution put forward in the project proposals.

We applied Algorithms 1 and 2 to the project data. In the evaluation, *Budget* is the only input variable, while we consider 5 output variables corresponding to different aspects of intended contribution from the project. Figure 1 displays the bootstrap outputs

by boxplots. The scores derived from the DEA CCR model (1) and the CE method are also marked in the figures. From Figure 1 we can see that the bootstrap Project 17 shows substantial variations. The reason is that Project 17 is *extremely efficient* in the DEA evaluation, because its input/output pair cannot be represented as a convex combination of other remaining projects, and therefore it will be sensitive to data variations; see also Andersen and Petersen (1993) and Lovell and Rouse (2003) for a pertinent discussion on the super-efficiency DEA model[6]. This can be an importance drawback in practice. For instance, such extremely efficient DMUs may be considered less favorable if we see variations in efficiency as an important evaluation criterion. The concomitant change in the evaluation result with the addition of variance criterion, however, does not always make sense, if we notice that extremely efficient DMUs at least dominate all the other units in the deterministic sense.

Finally, we can also see that the CE scores do not dominate the associated DEA scores.

In Figure 1, we see that the DEA score dominates the RCE score; this domination arises from the definition of RCE scores (11). We also have an interesting observation that the median and mean of bootstrap RCE scores are close to the CE scores[7]. This also implies a fair degree of symmetry of the bootstrap distributions, as we can also see in Figure 1. So, based on the probabilistic structure of wights, bootstrap RCE scores have distributions that symmetrically encompass the CE scores. Also RCE scores are in general less susceptible to the finite sample error, as compared to the CE scores (see Figure 1; and the discussion in the previous paragraph).

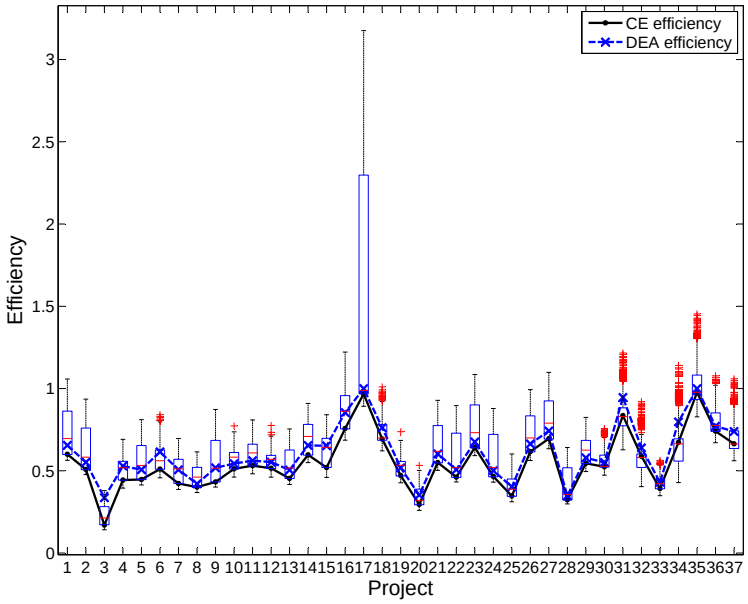
So far we have introduced two bootstrapping algorithms based on the CE method. However, the results of the algorithms do not automatically reveal how the information can be used for decision-making. Many managerial decisions concern the likelihood of certain events or variation revealed by distribution functions. Next, we present an application of our bootstrap algorithms, which use the mean-variance formulation to perform R&D project budgeting and supply chain configuration problem. In the next section, we will show how to select projects and allocate project budget based on the bootstrap outputs.

6 Application: R&D project selection and budgeting

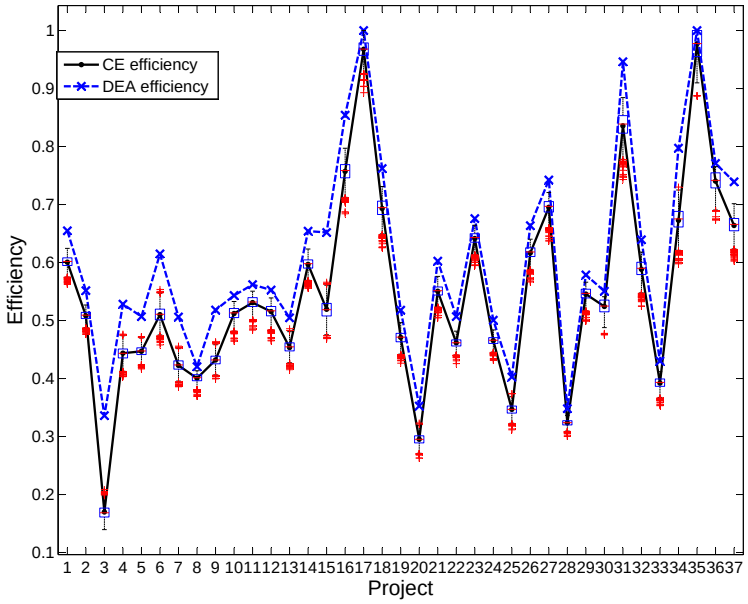
In project selection problems, decision-makers need to apportion a limited amount of resources to a set of projects, based on a number of evaluation criteria. The objective is to select the set of most "preferred" projects to finance. The evaluation criteria we

[6]In fact, STEP 6 of Algorithm 1 corresponds to a variant of the super-efficiency model.

[7]The average absolute deviation between the median and the CE score is 0.0136, while for the mean is 0.0011.



(a) Bootstrap outputs from Algorithm 1



(b) Bootstrap outputs from Algorithm 2

Figure 1: Empirical examinations of bootstrap distributions

often see are expected profits, time to market, and costs. The projects may be either independent or share their contributions and resources used. The example application considered in this section consists of the 37 independent projects that have been seen in the previous section. In addition to evaluating project performance, our job here is to allocate a fixed amount of resources across the candidate projects.

Product managers need to determine the new product portfolio and the related production planning before their competitors enter the market. The product portfolio decision is usually based on projected costs, demand, and expected profits within the product lifetime. Estimating these figures is particularly difficult for products that are new to the market and the company. To earn competitive advantages, firms across different sectors have to allocate limited resources to R&D projects according to the estimated project performance and required budget. Purchasing managers responsible for procuring new materials or products can only rely on limited data to select ideal suppliers. Decision makers in the above situations all need a systematic approach to evaluate product lines, products, or suppliers. Building an *a priori* probabilistic model about the risk involved can be difficult due to the lack of information, and erroneous decisions can give rise to huge financial loss. In face of the such uncertainty and substantial consequences, the evaluation process in these decision problems should be comprehensive, objective, and the results should be justifiable to all stakeholders, who may have their own viewpoints about the relative importance associated with each evaluation criterion.

Decision problems of this class has several typical characteristics: First, limited or no data, or no information are available. These dire situations are often coupled with limited or no data available for decision support, due to shorten product life cycles and market's pressing demand for more product innovation. Second, decisions concern about multiple performance indications of the process. Third, risk management.

6.1 Mean-variance formulation of portfolio optimization

In R&D project selections, huge investment is involved. Therefore it is crucial that managers consider the underlying risks before making the decision (see Huchzermeier and Loch, 2001). To incorporate risk factor into decision making, we have to know the probability distribution of the statistic of interest to evaluate the risk of decisions—however, the CE method cannot provide information about the distribution of CE scores, and there seems no obvious evidence that can lead us to the usual parametric structure.

The mean-variance formulation proposed by Markowitz (1952) has been a classic model in financial portfolio optimization. The model receives the name from it combination of the two most important factors, *return* and *risk*, by the *mean* and *variance* of the return distribution of an investment portfolio. Markowitz and Todd (2000) further conclude that the mean-variance model provides the maximum expected utility for most

most utility functions; see Wang and Xia (2002) for a detailed discussion. To construct a mean-variance model, however, we need information about the mean vector and variance-covariance matrix of the return distributions. The bootstrap methods developed in this paper allow us to approximate efficiency distributions of R&D projects, from which we can derive the estimated mean and variance of a project portfolio.

Consider an investor who receives n project proposals and has a fixed amount of resources to invest. The decision variables are the proportions of resources allotted to different projects, which is denoted by p_i, for $i = 1, \ldots, n$. We assume a budgetary rule under which the selected project has to be financed at least 70% of its requested budget. The vector $\mathbf{p} = [p_1, \ldots, p_n]'$ is called a portfolio. The efficiency of project k is a random variable θ_k with mean $\mathbb{E}(\theta_k)$. Denote Σ to be the variance-covariance matrix of all project efficiencies and $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_n]^T$. It then follows that the mean and variance of the efficiency of a portfolio $\mathbf{p}$ is

$$\mathbb{E}(\mathbf{p}'\boldsymbol{\theta}) = \mathbf{p}'\mathbb{E}(\boldsymbol{\theta}) \text{ and } Var(\mathbf{p}'\boldsymbol{\theta}) = \mathbf{p}'\Sigma\mathbf{p}, \text{ respectively.} \tag{16}$$

Suppose the investor wants the portfolio to be mean maximizing and variance minimizing. THen the portfolio selection can be naturally formulated as a bi-criterion problem, which reflects the trade-off between risk and return[8]. In practice we often scalarize the problem to a mix-integer quadratic problem (17), which is NP-hard (Jobst et al. 2001):

$$\begin{aligned}
\max \quad & \mathbf{p}'\mathbb{E}(\boldsymbol{\theta}) - \kappa(\mathbf{p}'\Sigma\mathbf{p}) && \text{(17a)}\\
\text{subject to} \quad & \mathbf{p}'\mathbf{1}_n = \delta && \text{(17b)}\\
& y_i l_i \delta \le p_i \le y_i u_i \delta, \ i = 1, \ldots, n && \text{(17c)}\\
& y_i \in \{0, 1\}, \ i = 1, \ldots, n. && \text{(17d)}
\end{aligned}$$

The parameter $\kappa \ge 0$ specifies the investor's degree of aversion to risk: the higher the value, the more risk-averse the solution will be. Constraint (17b) states that the portfolio ratios should sum up to one. The δ represents the budget available for allocation. Constraints (17c) and (17d) express the 70% budget rule: so l_i is set to 70% and $u_i = 1$. The binary variable y_i represent the choice whether project i is selected.

Based on the same mean-variance concept, we can also formulate an alternative model by rearranging (17) and adding a portfolio efficiency constraint:

[8]The mean maximizing (given a upper bound for variance) and variance minimizing (given a lower bound for return) portfolios are called mean-variance efficient portfolios.

Table 2: R&D project budgeting: a comparison[†]

Project	Green et al. (1996)	Liang et al. (2008b)	Algorithm 1	Algorithm 2
1	84.2	84.2	0.0	84.2(100%)
2	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0
5	0.0	0.0	0.0	0.0
6	0.0	0.0	0.0	0.0
7	0.0	0.0	0.0	0.0
8	0.0	0.0	0.0	0.0
9	0.0	0.0	0.0	0.0
10	0.0	0.0	54.3(70%)	0.0
11	0.0	0.0	53.5(70%)	76.5(100%)
12	47.5	47.5	47.5(100%)	47.5(100%)
13	0.0	0.0	58.5(100%)	58.5(100%)
14	95.0	95.0	0.0	0.0
15	0.0	83.8	0.0	0.0
16	35.4	35.4	35.4(100%)	35.4(100%)
17	32.1	32.1	32.1(100%)	32.1(100%)
18	46.7	46.7	46.7(100%)	46.7(100%)

[†] Percentage of the budget permitted.

$$\begin{aligned} \min \quad & \mathbf{p}'\Sigma\mathbf{p} && \text{(18a)} \\ \text{subject to} \quad & \mathbf{p}'\mathbf{1}_n = \delta && \text{(18b)} \\ & \mathbf{p}'\mathbb{E}(\boldsymbol{\theta}) \geq \gamma && \text{(18c)} \\ & y_i l_i \delta \leq p_i \leq y_i u_i \delta,\ i = 1, \ldots, n && \text{(18d)} \\ & y_i \in \{0, 1\},\ i = 1, \ldots, n. && \text{(18e)} \end{aligned}$$

In problem (18), γ in (18c) represents the lowest mean efficiency that the investor can accept.

To obtain estimates of the mean vector and variance-covariance matrix, we apply the bootstrap algorithms developed earlier to the project data (Tables 5). Here we only illustrate the application via problem (17); problem (18) can be adopted similarly. We are also given the constraint that the maximum budget for the R&D program cannot exceed 1000 monetary units. To solve the mean-variance model, we set $\kappa = 0.7$ in problem (17) and use the branch-and-bound algorithm combined with the quadratic programming solver of LINGO software to solve problem (17) to optimality.

Tables 2 and 3 list the project selection results in terms of the amount and the percentage of budget allocated permitted. The second and the third columns are the

Table 3: R&D project budgeting: a comparison (contd.)

Project	Green et al. (1996)	Liang et al. (2008b)	Algorithm 1	Algorithm 2
19	0.0	0.0	0.0	0.0
20	0.0	0.0	0.0	0.0
21	74.4	74.4	52.1(70%)	0.0
22	0.0	0.0	0.0	0.0
23	75.6	75.6	75.6(100%)	75.6(100%)
24	0.0	0.0	0.0	0.0
25	0.0	0.0	0.0	0.0
26	69.3	69.3	69.3(100%)	69.3(100%)
27	57.1	57.1	57.1(100%)	69.3(100%)
28	0.0	0.0	0.0	0.0
29	72	0.0	71.6(99.9%)	72.0(100%)
30	0.0	0.0	0.0	0.0
31	44.6	44.6	44.6(100%)	44.6(100%)
32	54.5	54.5	54.5(100%)	54.5(100%)
33	0.0	0.0	52.7(100%)	0.0
34	28.0	28.0	28.0(100%)	28.0(100%)
35	36.0	36.0	36.0(100%)	36.1(100%)
36	64.1	64.1	64.1(100%)	64.1(100%)
37	66.4	66.4	66.4(100%)	66.4(100%)

selection results determined by the alternative CE methods proposed in Green et al. (1996) and Liang et al. (2008b), respectively. The last two columns are the selection results from the bootstrap algorithms and the mean-variance model. By comparing the results from two categories of approaches (columns 2, 3 and columns 4, 5), we can first see the advantages of our methodology. First, although Green et al. (1996) and Liang et al. (2008b) have used alternative CE method to derive the CE scores, these two methods basically select projects based on the ordinal ranking of the efficiency scores, and therefore they are unable to incorporate "risks" into their selection process. Moreover, the previous approaches have to select each project in isolation and cannot consider the risk-mitigating effect of selecting several projects simultaneously, which is the cause of difference in the selection results between two sets of approaches (see Tables 2 and 3). Graves and Ringuest (2003) show that the former project selection approach will result in dominated project portfolios.

Second, in the former approached, decision-makers are forces to make binary selections if the projects based on the efficiency ranking without considering the integrated portfolio performance, while my approach can allocate the budget to maximize the mean-variance portfolio performance according to the variation within the efficiency estimation process.

7 Normalization procedures for bootstrap distributions

In this paper we have developed two bootstrap algorithms for the cross-efficiency estimators. In making decisions about the evaluated units, however, we still need some statistical procedures to elicit information from the bootstrap distributions. Many statistical procedures require a (multivariate) normality assumption on the populations; for example, ANOVA, discriminant analysis (see, e.g., Sharma, 1995), and many Ranking & Selection (R&S) methods and multiple comparisons methods (see, e.g., Kim and Nelson, 2006). The bootstrap distributions of efficiencies, however, are not guaranteed to be normally distributed. See, for example, the normality plots in Fig. 2, in which we can see the bootstrap distributions generated by our algorithms deviate from the hypothesized normal distributions. Using the Shapiro-Wilk normality test under 5% significance level shows these two distributions in the figure are significantly different from normal distributions ($p < 1\%$).

The non-normality poses a great problem when we intend to apply methods that require normality assumptions to the bootstrap outputs. In this section, we introduce a normalization procedure to preprocess the bootstrap data. In particular, we propose an algorithm based on the non-overlap batch means method to convert the simulation outputs to distributions of sample mean efficiencies. Formally, given a batch size k and a sequence of stochastic processes $\Theta_1, \Theta_2, \ldots, \Theta_m$ with $E(\Theta_i) = \theta$ and suppose that $m/k = n$ is an integer, we can divide the stochastic processes into n batches $\Theta(1)$ to $\Theta(n)$. The batch means is then defined by

$$\bar{\Theta}(i) = (\mathbf{1}_k' \Theta(i))/k, \text{ for } i = 1 \text{ to } n. \tag{19}$$

If k is large enough, it can be shown that $\mathbb{E}(\bar{\Theta}(i)) = \theta$ and the $\bar{\Theta}(i)$'s are uncorrelated (Law, 2007). We should note that the batch means method is mostly used to mitigate the influence of the *startup problem* in steady-state simulation experiments, for example, of queueing systems (Law, 2007). Kim and Nelson (2007) suggest using the batch means method to convert non-normal simulation outputs when normality is required.

In our case, we are concerned only about the normality condition of bootstrap distributions, rather than the dependence problem in most steady-state simulation experiments, since we use Monte Carlo simulation in the algorithms. Yet we still need to determine the batch size k, give a fixed number of bootstrap replications B. By the Central Limit Theorem, we know for sure that, as k increases to infinity, the batch means will be normally distributed. However, since we have a fixed number of B observations, specifying k too large will delete the samples of batched means. Therefore we can avoid this problem by searching for the minimal k that renders all batch means distributions normally

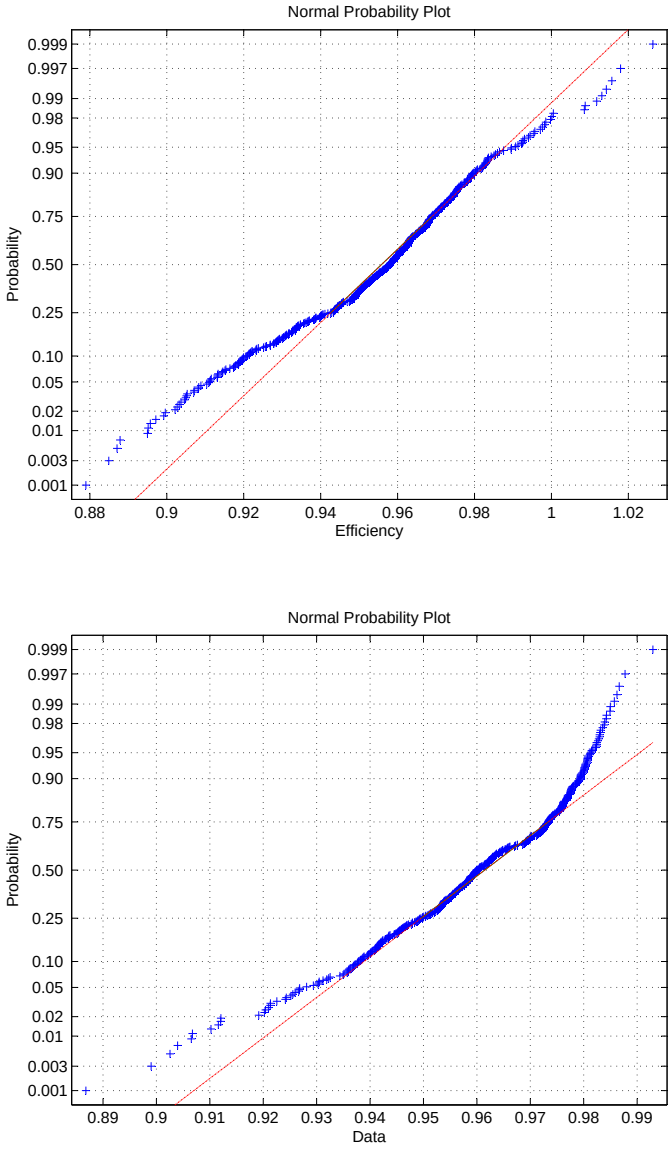


Figure 2: Bootstrap distributions of supplier-3's efficiency (Alg. 1 and 2, $B = 500$)

distributed. For univariate normality test, we apply the Shapiro-Wilk test (Shapiro and Wilk, 1965). Numerous tests exist for testing multivariate normality; for example, Royston's Multivariate Normality Test (Royston, 1983); see Mecklin and Mundfrom (2005) for a comparison of different testing procedures.

In determining the normality of distributions, however, we will perform multiple statistical testings. Therefore an appropriate procedure has to be applied to guard against the rising Type-I error probability due to multiple testings. One instant method to this problem is to make use of the Bonferroni inequality. Formally, let $E_1, E_2, \ldots, E_n$ denote n events in the sample space, the Bonferroni inequality given the following relationship:

$$P_r(\bigcap_{i=1}^{n} E_i^c) \leq \sum_{i=1}^{n} P_r(E_i^c). \tag{20}$$

Therefore (20) assures that if we conduct n comparisons at the significance level α/n, the overall Type-I error rate will be less than or equal to α. The Bonferroni method has several merits: it is easy to implement, and it can be applied regardless of the data structure. Unfortunately, one serious limitation of the method is its tendency to be overly conservative (i.e., the nominal overall significance level is much higher than the real value), and as a result the probability of Type II errors increases. We can also see that the problem will become serious as n grows larger. Other improved methods based on the Bonferroni's inequality also exist (see, e.g., Troendle, 1995).

In Bonferroni's procedure the overall significance level is divided by n because we can potentially make Type-I errors for n times. A simple way to improve is to account for the number of rejections in previous tests. So, for example, when we have already rejected k hypotheses, the significance level in the next test can be set to $\alpha/(n-k)$, while the overall Type-I error rate can still be bounded above by α. Hochberg's procedure is an extension of the idea Hochberg (1988), and we can easily prove that the procedure is uniformly more powerful than Bonferroni's. In Hochberg's procedure, we consider n null hypotheses $H_1, H_2, \ldots, H_n$ to be jointly tested. For hypothesis i, we can compute test statistics t_i and the associated p-value p_i, for $i = 1, \ldots, n$. Given an overall significance level α, Hochberg's procedure proceeds as follows:

Algorithm 3 Hochberg's procedure

1: sort p_i for $i = 1, \ldots, n$ to be $p_{[1]} \geq p_{[2]} \geq \ldots \geq p_{[n]}$. Set $i = 1$.
2: **if** $p_{[i]} \leq \alpha/i$ **then**
3: reject hypotheses from H_i to H_n and STOP.
4: **else**
5: accept H_i, increase i by one, and go to STEP 2.
6: **end if**

We should note that Hochberg's algorithm is constructed for a "static" situation. By

static I mean the sample sizes are predetermined and therefore the p-values are available beforehand. For the purpose of batch means, we need to adapt Hochberg's algorithm to accommodate the changing sample sizes.

Next we will propose an algorithm that deals with the issue of multiple hypothesis testings and progressively increase k until all distributions of batch means agree with the normality test. The steps are stated as follows:

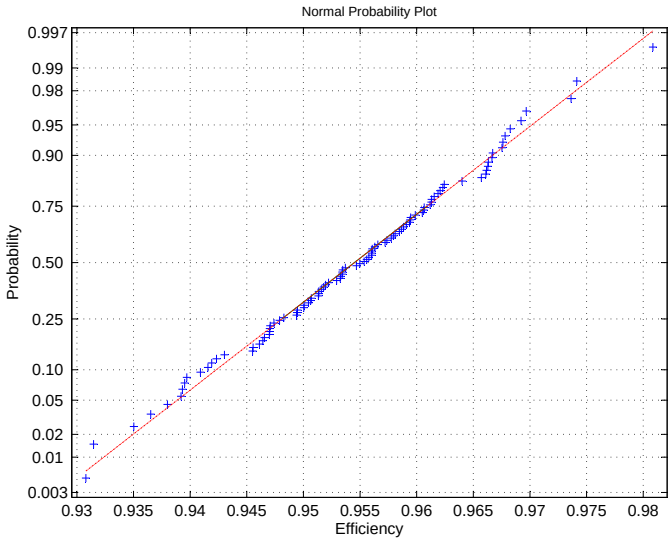
Algorithm 4 The batch means procedure

1: set the batch size $k = 1$, and the batch means sample size $n' = \lfloor n/k \rfloor$, and the nominal overall significance level α.
2: **for** $i = 1$ to n **do**
3: generate distributions of batch means θ_i based on k and n', for $i = 1$ to n.
4: calculate the Shapiro-Wilk test statistics W_i and the associated p-value p_i of bootstrap distribution for $i = 1, \ldots n$.
5: **end for**
6: call Algorithm 3.
7: **if** all hypotheses are maintained (i.e., $p_{[n]} \leq \alpha/n$) **then**
8: STOP.
9: **else**
10: increase the batch size by one, and return to STEP 2.
11: **end if**

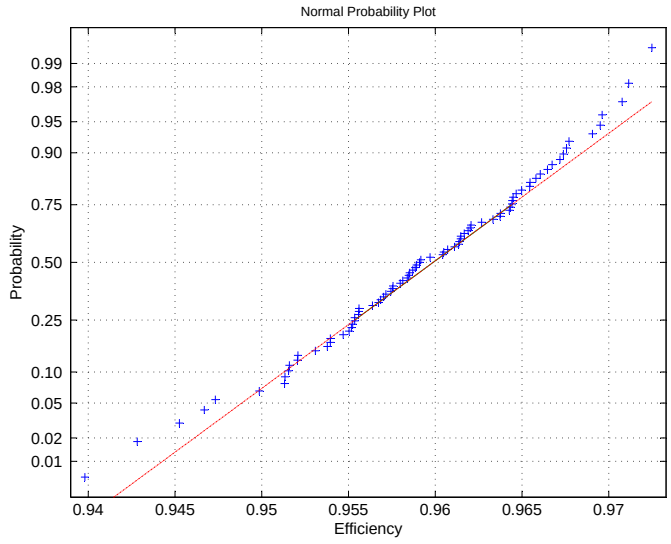
We apply Alg. 4 to the bootstrap distributions shown in Fig. 2. The algorithm terminates at the batch sizes of 5 and 6 for two distributions (with the replication size $B = 500$), respectively. We can also see from Fig. 3 that the batch means distributions are distributed closer to the hypothesized normal distribution, as compared to the original bootstrap distributions shown in Fig. 2.

A problem of Alg. 4 is that we cannot guarantee that that B is sufficient for achieving the normality condition, as the batch size also depends on how "non-normal" the bootstrap distributions are. In other words, we cannot determine a priori how large B should be. This issue, of course, can be solved by inserting an interactive procedure in Alg. 4, such that when the batch size increases to a certain extent we will need to do more bootstrap replications. In our case, though, setting $B = 500$ is sufficient for the procedure to stop before the batch size overflows the number of bootstrap replications (i.e., the sample sizes fall below 3). However, if we can increase B when necessary, the algorithm will always terminate in finite time by Central Limit Theorem and the principle of bootstrapping.

Lemma 1. *Alg. 4 will terminate in finite time given that B tends to infinity.*



(a) Batch means ($k = 5$) of the bootstrap distribution generated by Alg. 1



(b) Batch means ($k = 6$) of the bootstrap distribution generated by Alg. 2

Figure 3: Supplier-3's batch means of efficiency ($B = 500$)

7.1 Application: Supplier segregation models

Firms nowadays pay careful attention to the relationship with other firms in the supply network. By doing so, firms can stay closer to customers voice to reduce the uncertainty of demands. By working closely with suppliers, firms can also reduce product development costs, and improve the quality, and time-to-market of their new products (Handfield et al., 1999). The close bond between the buying firms and their suppliers makes it beneficial to improve and manage suppliers of products of strategic importance to create a win-win situation. With successful supplier management programs, both the buying firm's and the suppliers' strategic development road maps can be tightly coordinated (Krause et al., 1998; Krause, 1999).

Managing suppliers, however, requires that the buying firm has the insight of its suppliers performance or even can segregate them into different performance groups. Based on supplier's rating, the nature of the purchased items (e.g., strategic or non-strategic) and other strategic considerations, managers then have to do the cost-benefit analysis to decide whether to develop certain suppliers, or it is more beneficial to search for new suppliers.

In this final application of this paper, we illustrate how our developed algorithms—efficiency bootstrapping and batch means method—can be used to generate a supplier rating system. The rating, for example, is essential to the supplier development program. The buying firm can choose to interact with suppliers in different segregation clusters by different sourcing strategies and types of supplier relationships. As in the previous application, we consider the situation where we are given one observation (or estimation) of suppliers' inputs and outputs, either because the product or service is new to the market, or simply the historical information is not (fully) accessible because of the supplies' proprietary concern.

Fig. 4 shows the results of applying Alg. 2 and 4 to the supplier's performance data shown in Table 4, with the numbers indicating the class of suppliers. The classes of suppliers are determined by the simultaneous confidence intervals (SCIs) generated by Tukey's method for all pairwise comparisons[9]. The results indicate that Suppliers 3, 6, 13, 15 and 18 belong to the first class suppliers, while the second suppliers include suppliers 4 and 14, and so on. Based on the rating, the buying firm can formulate its procurement policies with respect to suppliers in different performance groups. For example, the buying firm can choose to do business only with the suppliers of the first two classes. Or it can develop a long-term partnership with those superior suppliers in the first two tiers, while foster competition for the under-performing suppliers. More development efforts can be given to, for example, the second-tier suppliers to encourage them to upgrade to be the first tier supplier so they can enjoy more benefits (e.g., balk

[9]Other approaches for multiple comparisons also exit; see, e.g., Hsu (1996).

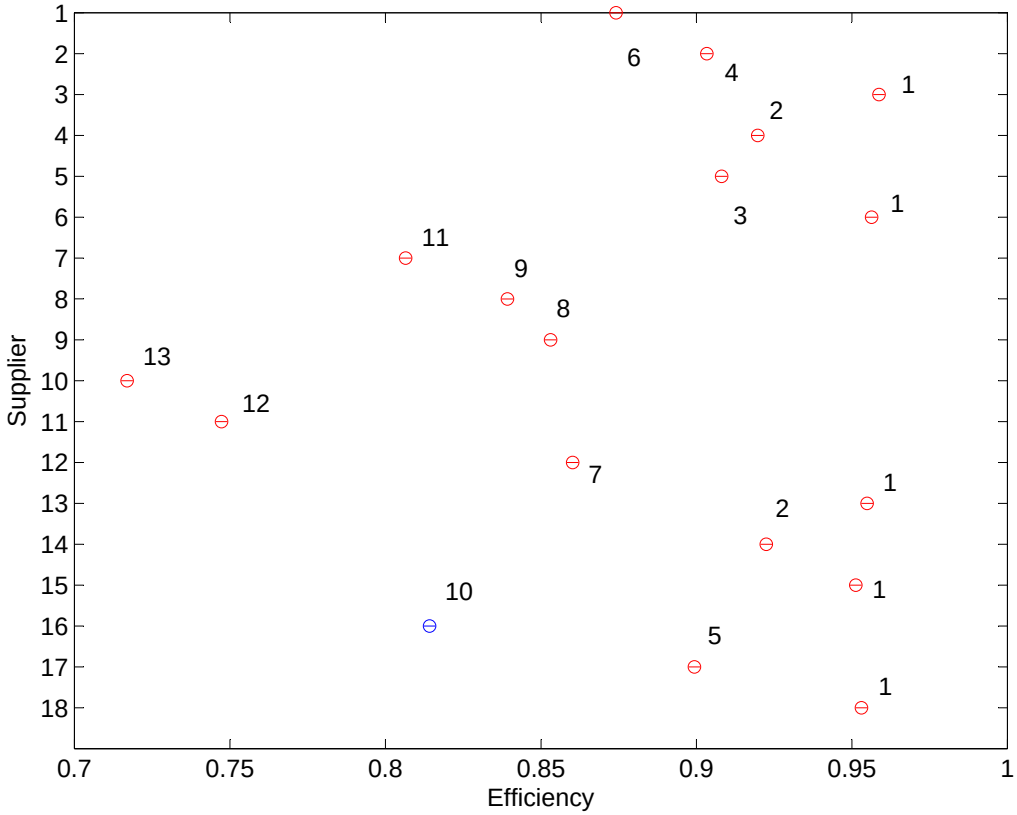


Figure 4: Simultaneous confidence intervals of supplier's efficiency based on Alg. 1

buying volume, long-term contracts).

8 Conclusions

I develop bootstrap algorithms based on the cross-efficiency method in DEA to evaluate policies, as well as decision-making methods for policy selection. Specifically, I formulate the algorithm based on the original CE definition and a revised CE index. The revised model rectify the sensitivity results that can arise in both the CE and DEA bootstrap outputs. One promising follow-up direction is to further investigate the relationship between CE and RCE methods. Although in the chapter we obtain some numerical results that indicate a close relationship between the mean/median of the bootstrap RCE distributions and the CE scores, we have yet to pinpoint the analytical connection between these two methods. Revealing this linkage can help the user further interpret the bootstrap results and its implication for the CE scores.

I also illustrate that we can use the bootstrap outputs to evaluate R&D projects and allocate resources according to the evaluation results. The proposed normalization procedure makes bootstrap outputs amenable to various statistical methods that require

a normality assumption.

Appendix: example data

Table 4: Supplier performance data set[†]

Supplier	No. of bills	No. on time	Experience	Credence	Total costs	No. shipped
1	90	187	240	90	253	197
2	130	194	210	80	268	198
3	200	220	270	70	259	229
4	100	160	200	70	180	169
5	173	204	160	70	257	212
6	170	192	230	80	248	197
7	60	194	200	90	272	209
8	145	195	170	60	330	203
9	150	200	180	70	327	208
10	90	171	170	60	330	203
11	100	174	200	80	321	207
12	200	209	210	100	329	234
13	163	165	330	90	281	173
14	170	199	250	80	309	203
15	185	188	250	90	291	193
16	85	168	240	80	334	177
17	130	177	210	70	249	185
18	160	167	200	80	216	176

[†] Source: Kleinsorge et al. (1992).

Table 5: R&D project proposal data set[†]

Project	Econ. contri.	Econ. contri.	Tech. contri.	Social contri.	Sci. contri.	Budget
1	67.53	70.82	62.64	44.91	46.28	84.20
2	58.94	62.86	57.47	42.84	45.64	90.00
3	22.27	9.68	6.73	10.99	5.92	50.20
4	47.32	47.05	21.75	20.82	19.64	67.50
5	48.96	48.48	34.90	32.73	26.21	75.40
6	58.88	77.16	35.42	29.11	26.08	90.00
7	50.10	58.20	36.12	32.46	18.90	87.40
8	47.46	49.54	46.89	24.54	36.35	88.80
9	55.26	61.09	38.93	47.71	29.47	95.90
10	52.40	55.09	53.45	19.52	46.57	77.50
11	55.13	55.54	55.13	23.36	46.31	76.50
12	32.09	64.04	33.57	40.60	29.36	47.50
13	27.49	39.00	34.51	21.25	25.74	58.50
14	77.17	83.35	60.01	41.37	51.91	95.00
15	72.00	68.32	25.84	36.64	25.84	83.80
16	39.74	34.54	38.01	15.79	33.06	35.40
17	38.50	28.65	51.18	59.59	48.82	32.10
18	41.23	47.18	41.01	10.18	38.86	46.70
19	53.02	51.34	42.48	17.42	46.30	78.60
20	19.91	18.98	25.49	8.66	27.04	54.10
21	50.96	53.56	55.47	30.23	50.44	82.10
22	53.36	46.47	49.72	36.53	50.44	82.10
23	61.60	66.59	64.54	39.10	51.12	75.60
24	52.56	55.11	57.58	39.69	56.49	92.30
25	31.22	29.84	33.08	13.27	36.75	68.50
26	54.64	58.05	60.03	31.16	46.71	69.30
27	50.40	53.58	53.06	26.68	48.85	57.10
28	30.76	32.45	36.63	25.45	34.79	80.00
29	48.97	54.97	51.52	23.02	45.75	72.00
30	59.68	63.78	54.80	15.94	44.04	82.90
31	48.28	55.58	53.30	7.61	36.74	44.60
32	39.78	51.69	35.10	5.30	29.57	54.50
33	24.93	29.72	28.72	8.38	23.45	52.70
34	22.32	33.12	18.94	4.03	9.58	28.00
35	48.83	53.41	40.82	10.45	33.72	36.00
36	61.45	70.22	58.26	19.53	49.33	64.10
37	57.78	72.10	43.83	16.14	31.32	66.40

[†] Indirect economic contribution, Direct economic contribution, Technical contribution, Social contribution, Scientific contribution, Required budget. Source: Oral et al. (1991).

References

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.

Andersen, P., Petersen, N., 1993. A procedure for ranking efficient units in data envelopment analysis. Management Science 39 (10), 1261–1264.

Banker, R., 1993. Maximum likelihood, consistency and Data Envelopment Analysis: A statistical foundation. Management Science 39 (10), 1265–1273.

Banker, R., Chang, H., 1995. A simulation study of hypothesis tests for differences in efficiencies. International Journal of Production Economics 39 (1-2), 37–54.

Banker, R., Charnes, A., Cooper, W., 1984. Some models for estimating technical and scale inefficiencies in Data Envelopment Analysis. Management Science 30, 1078–1092.

Ben-Tal, A., Nemirovski, A., 1999. Robust solutions of uncertain linear programs. Operations Research Letters 25 (1), 1–13.

Bertsimas, D., Thiele, A., 2004. A robust optimization approach to supply chain management. Lecture Notes in Computer Science, 86–100.

Bourgeois, I. L., Eisenhardt, K., 1988. Strategic decision processes in high velocity environments: four cases in the microcomputer industry. Management Science 34 (7), 816–835.

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

Chen, T., 2002. An assessment of technical efficiency and cross-efficiency in Taiwan's electricity distribution sector. European Journal of Operational Research 137 (2), 421–433.

Coelli, T., 2005. An introduction to efficiency and productivity analysis. Springer.

Cooper, W., Tone, K., Seiford, L., 2006. Introduction to Data Envelopment Analysis and its uses: with DEA-solver software and references. Springer.

Despotis, D., 2002. Improving the discriminating power of DEA: focus on globally efficient units. Journal of Operational Research Society 53, 314–325.

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.

Dyson, R., Allen, R., Camanho, A., Podinovski, V., Sarrico, C., Shale, E., 2001. Pitfalls and protocols in DEA. European Journal of Operational Research 132 (2), 245–259.

Efron, B., 1987. Better bootstrap confidence intervals. Journal of the American Statistical Association 82 (397), 171–185.

Efron, B., Tibshirani, R., 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Statistical Science 1 (1), 54–77.

Efron, B., Tibshirani, R., 1993. An introduction to the bootstrap. Chapman & Hall/CRC.

Graves, S., Ringuest, J., 2003. Models & methods for project selection: concepts from management science, finance and information technology. Springer.

Green, R., Doyle, J., Cook, W., 1996. Preference voting and project ranking using DEA and cross-evaluation. European Journal of Operational Research 90 (3), 461–472.

Hall, P., 1992. The bootstrap and Edgeworth expansion. Springer Verlag.

Handfield, R., Ragatz, G., Peterson, K., Monczka, R., 1999. Involving suppliers in new product development? California Management Review 42 (1), 60.

Hochberg, Y., 1988. A sharper Bonferroni procedure for multiple tests of significance. Biometrika 75 (4), 800–802.

Hsu, J., 1996. Multiple comparisons: theory and methods. Chapman & Hall/CRC.

Huchzermeier, A., Loch, C., 2001. Project management under risk: using the real options approach to evaluate flexibility in R&D. Management Science 47 (1), 85–101.

Jobst, N., Horniman, M., Lucas, C., Mitra, G., 2001. Computational aspects of alternative portfolio selection models in the presence of discrete asset choice constraints. Quantitative Finance 1 (5), 489–501.

Kim, S., Nelson, B., 2006. Selecting the best system. Handbooks in operations research and management science: Simulation.

Kim, S., Nelson, B., 2007. Recent advances in ranking and selection. Proceedings of the 39th conference on Winter simulation, 162–172.

Kleinsorge, I., Schary, P., Tanner, R., 1992. Data Envelopment Analysis for monitoring customer-supplier relationships. Journal of Accounting and Public Policy 11 (4), 357–372.

Kneip, A., Park, B., Simar, L., 1998. A note on the convergence of nonparametric DEA estimators for production efficiency scores. Econometric Theory 14 (06), 783–793.

Korostelev, A., Simar, L., Tsybakov, A., 1995. Efficient estimation of monotone boundaries. The Annals of Statistics 23 (2), 476–489.

Krause, D., 1999. The antecedents of buying firms' efforts to improve suppliers. Journal of Operations Management 17 (2), 205–224.

Krause, D., Handfield, R., Scannell, T., 1998. An empirical investigation of supplier development: reactive and strategic processes. Journal of Operations Management 17 (1), 39–58.

Law, A., 2007. Simulation modeling and analysis, 4th Edition. McGraw-Hill.

Liang, L., Wu, J., Cook, W., Zhu, J., 2008a. Alternative secondary goals in DEA cross-efficiency evaluation. International Journal of Production Economics.

Liang, L., Wu, J., Cook, W., Zhu, J., 2008b. The DEA cross efficiency model and its Nash equilibrium. Operations Research, forthcoming.

Lovell, C., Rouse, A., 2003. Equivalent standard DEA models to provide super-efficiency scores. Journal of the Operational Research Society 54 (1), 101–108.

Markowitz, H., 1952. Portfolio selection. Journal of Finance 7 (1), 77–91.

Markowitz, H., Todd, G., 2000. Mean-Variance analysis in portfolio choice and capital markets. Wiley.

Mecklin, C., Mundfrom, D., 2005. A Monte Carlo comparison of the Type I and Type II error rates of tests of multivariate normality. Journal of Statistical Computation and Simulation 75 (2), 93–107.

Oral, M., Kettani, O., Lang, P., 1991. A methodology for collective evaluation and selection of industrial R&D projects. Management Science 37 (7), 871–885.

Petrin, A., 2002. Quantifying the benefits of new products: the case of the minivan. Journal of Political Economy 110 (4), 705–729.

Royston, J., 1983. Some techniques for assessing multivariate normality based on the Shapiro-Wilk W. Applied Statistics 32 (2), 121–133.

Sexton, T., Silkman, R., Hogan, A., 1986. Data envelopment analysis: Critique and extensions. New Directions for Program Evaluation 1986 (32), 73–105.

Shang, J., Sueyoshi, T., 1995. A unified framework for the selection of a Flexible Manufacturing System. European Journal of Operational Research 85 (2), 297–315.

Shapiro, S., Wilk, M., 1965. An analysis of variance test for normality. Biometrika 52 (3), 591–599.

Sharma, S., 1995. Applied multivariate techniques. John Wiley & Sons, Inc., NY, USA.

Simar, L., Wilson, P., 1998. Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. Management Science 44 (1), 49–61.

Simar, L., Wilson, P., 1999. Estimating and bootstrapping Malmquist indices. European Journal of Operational Research 115 (3), 459–471.

Simar, L., Wilson, P., 2000a. A general methodology for bootstrapping in non-parametric frontier models. Journal of Applied Statistics 27 (6), 779–802.

Simar, L., Wilson, P., 2000b. Statistical inference in nonparametric frontier models: the state of the art. Journal of Productivity Analysis 13 (1), 49–78.

Talluri, S., Narasimhan, R., 2004. A methodology for strategic sourcing. European Journal of Operational Research 154 (1), 236–250.

Troendle, J., 1995. A stepwise resampling method of multiple hypothesis testing. Journal of the American Statistical Association 90 (429), 370–378.

Wang, S., Xia, Y., 2002. Portfolio selection and asset pricing. Springer.

Zhu, J., 2001. Super-efficiency and DEA sensitivity analysis. European Journal of Operational Research 129 (2), 443–455.

Publications in the Report Series Research* in Management

ERIM Research Program: "Business Processes, Logistics and Information Systems"

2008

An Evolutionary Framework for Determining Heterogeneous Strategies in Multi-Agent Marketplaces
Alexander Babanov, Wolfgang Ketter and Maria Gini
ERS-2008-002-LIS
http://hdl.handle.net/1765/10972

Choosing between Auctions and Negotiations in Online B2B Markets for IT Services: The Effect of Prior Relationships and Performance
Uladzimir Radkevitch, Eric van Heck and Otto Koppius
ERS-2008-004-LIS
http://hdl.handle.net/1765/11288

Key Issues in Expansion of End-User Mobile Communication in China
Sunanda Sangwan, Guan Chong and Louis-Francois Pau
ERS-2008-011-LIS
http://hdl.handle.net/1765/11762

Some Comments on the Question Whether Co-Occurrence Data Should Be Normalized
Ludo Waltman and Nees Jan van Eck
ERS-2008-014-LIS
http://hdl.handle.net/1765/11810

Bibliometric Mapping of the Computational Intelligence Field
Nees Jan van Eck and Ludo Waltman
ERS-2008-015-LIS
http://hdl.handle.net/1765/11811

Demand Management Opportunities in E-fulfillment: What Internet Retailers Can Learn from Revenue Management
Niels Agatz, Ann Campbell, Moritz Fleischmann, Jo van Nunen and Martin Savelsbergh
ERS-2008-021-LIS
http://hdl.handle.net/1765/12244

Time Slot Management in Attended Home Delivery
Niels Agatz, Ann Campbell, Moritz Fleischmann and Martin Savelsbergh
ERS-2008-022-LIS
http://hdl.handle.net/1765/12245

Mobile Service Affordability for the Needy, Addiction, and ICT Policy Implications
L-F Pau
ERS-2008-023-LIS
http://hdl.handle.net/1765/12246

CBPRS: A City Based Parking and Routing System
J.L. Boehlé, L.J.M. Rothkrantz and M. van Wezel
ERS-2008-029-LIS
http://hdl.handle.net/1765/12467

Coping with Costly Bid Evaluation in Online Reverse Auctions for IT Services
Uladzimir Radkevitch, Eric van Heck and Otto Koppius
ERS-2008-039-LIS
http://hdl.handle.net/1765/12871

Multi Agent Systems in Logistics: A Literature and State-of-the-art Review
Niels Lang, Hans M. Moonen, F. Jordan Srour and Rob A. Zuidwijk
ERS-2008-043-LIS
http://hdl.handle.net/1765/12902

Generalizing the h- and g-indices
Nees Jan van Eck and Ludo Waltman
ERS-2008-049-LIS
http://hdl.handle.net/1765/13210

Applying Revenue Management to the Reverse Supply Chain
Mark Ferguson, Moritz Fleischmann and Gilvan C. Souza
ERS-2008-052-LIS
http://hdl.handle.net/1765/13211

Ergonomics Contributions to Company Strategies
Jan Dul and W. Patrick Neumann
ERS-2008-058-LIS
http://hdl.handle.net/1765/13213

Can the Office Environment Stimulate a Manager's Creativity?
Canan Ceylan, Jan Dul and Serpil Aytac
ERS-2008-059-LIS
http://hdl.handle.net/1765/13215

Tactical and Strategic Sales Management for Intelligent Agents Guided By Economic Regimes
Wolfgang Ketter, John Collins, Maria Gini, Alok Gupta, and Paul Schrater
ERS-2008-061-LIS
http://hdl.handle.net/1765/13547

How Much is Location Information Worth? A Competitive Analysis of the Online Traveling Salesman Problem with Two Disclosure Dates
F. Jordan Srour and Rob Zuidwijk
ERS-2008-075-LIS
http://hdl.handle.net/1765/13837

Automatic Term Identification for Bibliometric Mapping
Nees Jan van Eck, Ludo Waltman, Ed C.M. Noyons, and Reindert K. Buter
ERS-2008-081-LIS
http://hdl.handle.net/1765/14056

Flexible Decision Support in Dynamic Interorganizational Networks
John Collins, Wolfgang Ketter, and Maria Gini
ERS-2008-082-LIS
http://hdl.handle.net/1765/14082

Multi-Factor Policy Evaluation and Selection in the One-Sample Situation
Chien-Ming Chen
ERS-2008-084-LIS
http://hdl.handle.net/1765/14275

---

* A complete overview of the ERIM Report Series Research in Management:
https://ep.eur.nl/handle/1765/1

ERIM Research Programs:
LIS Business Processes, Logistics and Information Systems
ORG Organizing for Performance
MKT Marketing
F&A Finance and Accounting
STR Strategy and Entrepreneurship