European Journal of Operational Research 218 (2012) 186–192

Contents lists available at SciVerse ScienceDirect

European Journal of Operational Research

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

Decision Support

Additive super-efficiency in integer-valued data envelopment analysis

Juan Du [a], Chien-Ming Chen [b], Yao Chen [c], Wade D. Cook [d], Joe Zhu [e,*]

[a] School of Economics and Management, Tongji University, 1239 Siping Road, Shanghai 200092, PR China
[b] Nanyang Business School, Nanyang Technological University, Singapore 639798, Singapore
[c] College of Management, University of Massachusetts at Lowell, Lowell, MA 01845, USA
[d] Schulich School of Business, York University, Toronto, Ontario, Canada M3J 1P3
[e] School of Business, Worcester Polytechnic Institute, Worcester, MA 01609, USA

ARTICLE INFO

Article history:
Received 3 January 2011
Accepted 12 October 2011
Available online 4 November 2011

Keywords:
Data envelopment analysis
Early-stage ventures
Integer-valued data
Slacks
Super-efficiency

ABSTRACT

Conventional data envelopment analysis (DEA) methods assume that input and output variables are continuous. However, in many real managerial cases, some inputs and/or outputs can only take integer values. Simply rounding the performance targets to the nearest integers can lead to misleading solutions and efficiency evaluation. Addressing this kind of integer-valued data, the current paper proposes models that deal directly with slacks to calculate efficiency and super-efficiency scores when integer values are present. Compared with standard radial models, additive (super-efficiency) models demonstrate higher discrimination power among decision making units, especially for integer-valued data. We use an empirical application in early-stage ventures to illustrate our approach.

1. Introduction

Developed by Charnes et al. (1978), data envelopment analysis (DEA) is an effective approach to measuring the relative efficiency of peer decision making units (DMUs) with multiple inputs and outputs. Standard DEA models assume real values for all inputs and outputs. However, in many real-world situations, some inputs and/or outputs can only take integer values. For example, when applying DEA method to analyze the efficiency for hospitals, inputs such as the number of doctors and outputs such as the number of completed surgeries are restricted to non-negative integers. As pointed out by Kuosmanen and Kazemi Matin (2009), simply rounding the optimal solution to the nearest whole numbers can lead to misleading efficiency evaluations and performance targets in some cases, especially for those small-size DMUs with small input and output scales. Consider a similar example from Kuosmanen and Kazemi Matin (2009). Suppose a local hospital has 4 full-time physicians, and the DEA analysis suggests that the efficient benchmark for full-time physicians is 3.5. Rounding the reference figure down to 3 may lead to a shortage in clinical providers to meet the service standards, while rounding the target up to 4 does not save any resources even though the hospital is inefficient in DEA.

Lozano and Villa (2006) address the integer-valued issue in DEA by proposing a mixed integer linear programming (MILP) model that restricted the computed targets to integers. However, as later argued by Kuosmanen and Kazemi Matin (2009), this paper has the following two issues. One is that the theoretical foundation of the model is ambiguous. It does not comply with the minimum extrapolation principle (Banker et al., 1984), which is the foundation of all DEA models. The other is that Lozano and Villa's (2006) MILP model tends to overestimate the efficiency results. To tackle both of the above problems, Kuosmanen and Kazemi Matin (2009) develop a new axiomatic foundation for integer-valued DEA models, and further derive a new production possibility set (PPS) for integer-valued DEA with constant returns to scale (CRS) by introducing new axioms of *natural disposability* and *natural divisibility*. Based upon the new axiomatic foundation and PPS, they propose an improved version of MILP formulation for efficiency assessment that avoids overestimations of efficiency scores. According to Kuosmanen and Kazemi Matin (2009), the terms of *natural disposability* and *natural divisibility* are integer-valued versions of the conventional *free disposability* and *non-increasing returns to scale* axioms, respectively. Later, Kazemi Matin and Kuosmanen (2009) extend the axiomatic foundation for integer-valued DEA with CRS in their previous work (Kuosmanen and Kazemi Matin, 2009) to other situations like variable, non-decreasing and non-increasing returns to scale. To achieve such a generalization, they introduce some new axioms. For example, under the assumption of variable returns to scale (VRS), a new notion of *natural convexity*

* Corresponding author. Tel.: +1 508 831 5467; fax: +1 508 831 5720.
E-mail addresses: dujuan@tongji.edu.cn (J. Du), cmchen@ntu.edu.sg (C.-M. Chen), Yao_Chen@uml.edu (Y. Chen), wcook@schulich.yorku.ca (W.D. Cook), jzhu@wpi.edu (J. Zhu).

0377-2217/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejor.2011.10.023

is introduced which restricts the feasible convex combinations only to the subset made up of integer-valued points.

The afore-mentioned literature all use standard radial DEA models to analyze efficiency for integer-valued data. In our empirical study, we discover that these existing DEA approaches are often not able to appropriately discriminate among the performance measures of DMUs. In this paper we take a different approach. We extend the efficiency analysis for integer-valued data by dealing directly with input and output slacks, and develop additive integer-valued efficiency and super-efficiency models. In doing so, we also extend the slacks-based measures developed by Tone (2001, 2002). For an efficient DMU, its super-efficiency score is computed based on the related input and output slacks. This slacks-based efficiency analysis demonstrates a stronger discriminating power among DMUs, especially for those classified as efficient. This is a useful feature in providing decision makers with insights into the performance of peer DMUs, and in helping carry out further analysis for managerial decisions. Later, we will illustrate our approach and models through an empirical study in the financial success of new projects.

This empirical study uses the data collected by the Inventor's Assistance Program (IAP) at the Canadian Innovation Center (CIC) in Waterloo, Canada. We use the same data set from IAP with a different research perspective from previous literatures such as Åstebro and Elhedhli (2006). Since evaluating scores provided by the analysts are integers, we apply our integer-valued efficiency and super-efficiency analysis to further explore and compare the potential performance of early-stage ventures. The results illustrate that efficient projects have a much better chance in financial success compared with the entire project pool or inefficient projects. In the application, we compare the additive efficiency and super-efficiency with their radial counterparts, and find that efficient DMUs can be differentiated more explicitly by additive super-efficiency than by radial super-efficiency which does not consider slacks.

The rest of this paper is organized as follows. Section 2 proposes radial and additive DEA models for integer-valued data set. To further distinguish those efficient DMUs, Section 3 presents radial and additive super-efficiency DEA models within an integer context. Section 4 discusses an empirical study in new business development, with the data collected by IAP at CIC. Section 5 concludes with a summary of our contributions.

2. Radial and additive integer-valued DEA models

2.1. Radial model

Assume that there are n DMUs producing the same set of outputs by consuming the same set of inputs. Unit j is denoted by DMU_j ($j = 1,\ldots,n$), whose ith input and rth output are denoted by x_{ij} ($i = 1,\ldots,m$) and y_{rj} ($r = 1,\ldots,s$), respectively. Then the input-oriented efficiency of DMU_o with variable returns to scale (VRS) is evaluated by the following linear program (1) (Banker et al., 1984):

$$\begin{aligned} \min \quad & \theta_o - \varepsilon\left(\sum_{i=1}^{m} s_{io}^{-} + \sum_{r=1}^{s} s_{ro}^{+}\right) \\ s.t. \quad & \sum_{j=1}^{n} \lambda_j x_{ij} + s_{io}^{-} = \theta_o x_{io}, \quad i = 1,\ldots,m \\ & \sum_{j=1}^{n} \lambda_j y_{rj} - s_{ro}^{+} = y_{ro}, \quad r = 1,\ldots,s \\ & \sum_{j=1}^{n} \lambda_j = 1 \\ & \lambda_j, s_{io}^{-}, s_{ro}^{+} \geqslant 0, \quad j = 1,\ldots,n,\ i = 1,\ldots,m,\ r = 1,\ldots,s \end{aligned} \tag{1}$$

where ε represents a non-Archimedean infinitesimal.

Suppose that some of the inputs and outputs are restricted to integer values. We follow the notations used by Kuosmanen and Kazemi Matin (2009), and denote the subsets of integer-valued and real-valued inputs, integer-valued and real-valued outputs by I^I, I^{NI}, O^I and O^{NI}, respectively. Obviously, I^I and I^{NI}, as well as O^I and O^{NI} are mutually disjoint, and satisfy $I^I \cup I^{NI} = \{1,2,\ldots,m\}$, $O^I \cup O^{NI} = \{1,2,\ldots,s\}$. By changing input-orientation into output-orientation, we obtain the output-oriented VRS version of Kuosmanen and Kazemi Matin's (2009) mixed integer linear programming (MILP) model (2) as follows. Here we choose to present the output-oriented VRS model because it suits better with the features of our application data set. This will be further explained in Section 4—the application part

$$\begin{aligned} Eff_o^O = \max \quad & \alpha_o + \varepsilon\left(\sum_{i\in I} s_{io}^{-} + \sum_{r_{NI}\in O^{NI}} s_{r_{NI}o}^{+} + \sum_{r_I\in O^I} s_{r_Io}^{+} + \sum_{r_I\in O^I} \tilde{s}_{r_Io}^{+}\right) \\ s.t. \quad & x_{io} - s_{io}^{-} = \sum_{j=1}^{n} \lambda_j x_{ij}, \quad i \in I \\ & \alpha_o y_{r_{NI}o} + s_{r_{NI}o}^{+} = \sum_{j=1}^{n} \lambda_j y_{r_{NI}j}, \quad r_{NI} \in O^{NI} \\ & \tilde{y}_{r_Io} + s_{r_Io}^{+} = \sum_{j=1}^{n} \lambda_j y_{r_Ij}, \quad r_I \in O^I \\ & \alpha_o y_{r_Io} + \tilde{s}_{r_Io}^{+} = \tilde{y}_{r_Io}, \quad r_I \in O^I \\ & \tilde{y}_{r_Io} \in Z^{+}, \quad r_I \in O^I \\ & \sum_{j=1}^{n} \lambda_j = 1 \\ & \lambda_j \geqslant 0, \quad j = 1,\ldots,n \\ & s_{io}^{-} \geqslant 0, \quad s_{r_{NI}o}^{+} \geqslant 0, \quad s_{r_Io}^{+} \geqslant 0, \quad \tilde{s}_{r_Io}^{+} \geqslant 0, \\ & i \in I, \quad r_{NI} \in O^{NI}, \quad r_I \in O^I \end{aligned} \tag{2}$$

where variables s_{io}^{-}, $s_{r_{NI}o}^{+}$, $s_{r_Io}^{+}$, $\tilde{s}_{r_Io}^{+}$ represent the non-radial slacks, α_o represents the output-oriented efficiency for DMU_o, while $\tilde{y}_{r_Io} \in Z^{+}$ is the integer-valued reference point for outputs O^I. DMU_o is considered efficient if the optimal value for α_o equals one. It is worth noting that the above model (2) distinguishes between two types of output slacks. The first type of slack variables, denoted by $s_{r_{NI}o}^{+}$ and $s_{r_Io}^{+}$, represents the absolute difference between the convex combination $\sum_{j=1}^{n}\lambda_j y_{r_{NI}j}$ (or $\sum_{j=1}^{n}\lambda_j y_{r_Ij}$) and the reference point $\alpha_o y_{r_{NI}o}$ (or $\tilde{y}_{r_Io}$). The second type denoted by $\tilde{s}_{r_Io}^{+}$, represents the absolute difference between the reference point $\tilde{y}_{r_Io}$ and the projection $\alpha_o y_{r_Io}$ for integer-restricted outputs. We note in particular that $\tilde{s}_{r_Io}^{+}$ and $s_{r_Io}^{+}$ are two totally different slack variables concerning output r_I. $\tilde{s}_{r_Io}^{+}$ is used to adjust the projection $\alpha_o y_{r_Io}$ to its integer reference point $\tilde{y}_{r_Io}$, while $s_{r_Io}^{+}$ measures the absolute distance between this integer reference target $\tilde{y}_{r_Io}$ and the convex combination $\sum_{j=1}^{n}\lambda_j y_{r_Ij}$, thus $\sum_{j=1}^{n}\lambda_j y_{r_Ij}$ is free from the integer restriction.

2.2. Additive models

Note that the above model (2) and its input-oriented version only impose integer restrictions either on outputs or inputs. We can incorporate both in one model by dealing directly with slacks. By adding the integer requirement in an additive DEA model (Charnes et al., 1982), we propose the following model (3), referred to as the additive integer-valued DEA model.

$$
\begin{aligned}
\hat{\rho}_o^* = \max \quad & \hat{\rho}_o = \sum_{i_{NI}\in I^{NI}} s^-_{i_{NI}o} + \sum_{r_{NI}\in O^{NI}} s^+_{r_{NI}o} + \sum_{i_I\in I^I} \tilde{s}^-_{i_Io} + \sum_{r_I\in O^I} \tilde{s}^+_{r_Io} \\
s.t. \quad & x_{i_{NI}o} - s^-_{i_{NI}o} = \sum_{j=1}^n \lambda_j x_{i_{NI}j}, \quad i_{NI}\in I^{NI} \\
& y_{r_{NI}o} + s^+_{r_{NI}o} = \sum_{j=1}^n \lambda_j y_{r_{NI}j}, \quad r_{NI}\in O^{NI} \\
& \tilde{x}_{i_Io} \geqslant \sum_{j=1}^n \lambda_j x_{i_Ij}, \quad i_I \in I^I \\
& x_{i_Io} - \tilde{s}^-_{i_Io} = \tilde{x}_{i_Io}, \quad i_I \in I^I \\
& \tilde{y}_{r_Io} \leqslant \sum_{j=1}^n \lambda_j y_{r_Ij}, \quad r_I \in O^I \\
& y_{r_Io} + \tilde{s}^+_{r_Io} = \tilde{y}_{r_Io}, \quad r_I \in O^I \\
& \sum_{j=1}^n \lambda_j = 1, \quad \lambda_j \geqslant 0,\ j = 1,\ldots,n \\
& \tilde{x}_{i_Io}, \tilde{y}_{r_Io} \in Z^+, \quad i_I \in I^I,\ r_I \in O^I \\
& s^-_{i_{NI}o}, s^+_{r_{NI}o}, \tilde{s}^-_{i_Io}, \tilde{s}^+_{r_Io} \geqslant 0, \quad i_{NI}\in I^{NI},\ r_{NI}\in O^{NI}, \\
& i_I \in I^I,\ r_I \in O^I
\end{aligned} \tag{3}
$$

In model (3), $\tilde{x}_{i_Io} \in Z^+$ and $\tilde{y}_{r_Io} \in Z^+$ are the integer-valued targets for input i_I and output r_I of DMU$_o$. Non-radial slacks $s^-_{i_{NI}o}$, $s^+_{r_{NI}o}$, $\tilde{s}^-_{i_Io}$, $\tilde{s}^+_{r_Io}$ represent the actual inputs that can be reduced and actual outputs that can be increased in order to realize the best feasible target.

Note that as in the standard additive DEA model, model (3) does not offer an integrated efficiency score between zero and one. One can, however, define a posteriori efficiency index based upon a set of optimal solution from model (3). For example, according to the slacks-based measure of (SBM) efficiency proposed in Tone (2001, 2002), we can similarly define $\hat{\sigma}_o^* = \frac{1-\frac{1}{m}\left[\sum_{i_{NI}\in I^{NI}} s^{-*}_{i_{NI}o}/x_{i_{NI}o} + \sum_{i_I\in I^I} \tilde{s}^{-*}_{i_Io}/x_{i_Io}\right]}{1+\frac{1}{s}\left[\sum_{r_{NI}\in O^{NI}} s^{+*}_{r_{NI}o}/y_{r_{NI}o} + \sum_{r_I\in O^I} \tilde{s}^{+*}_{r_Io}/y_{r_Io}\right]}$ as the *additive efficiency* measure for DMU$_o$, where $\left\{\lambda_j^*,\ j = 1,\ldots, n;\ s^{-*}_{i_{NI}o},\ i_{NI}\in I^{NI};\ s^{+*}_{r_{NI}o},\ r_{NI}\in O^{NI};\ \tilde{s}^{-*}_{i_Io},\ i_I\in I^I;\ \tilde{s}^{+*}_{r_Io},\ r_I\in O^I\right\}$ is an optimal solution to model (3). Here the additive efficiency score is computed based upon an optimal solution of the additive model (3). In the DEA literature, similar posterior efficiency indices have been developed/defined in an attempt to incorporate slacks after radial DEA score is obtained. See, for example, Torgersen et al. (1996), and Chen and Sherman (2004).

In order to make the resulting model unit-invariant, alternative objective functions can also be used for model (3). One possible choice could be

$$
\hat{\eta}_o^* = \max \hat{\eta}_o = \frac{1}{m+s}\left[\sum_{i_{NI}\in I^{NI}} \frac{s^-_{i_{NI}o}}{x_{i_{NI}o}} + \sum_{r_{NI}\in O^{NI}} \frac{s^+_{r_{NI}o}}{y_{r_{NI}o}} + \sum_{i_I\in I^I} \frac{\tilde{s}^-_{i_Io}}{x_{i_Io}} + \sum_{r_I\in O^I} \frac{\tilde{s}^+_{r_Io}}{y_{r_Io}}\right] \tag{4}
$$

The constraints from model (3) indicate that $x_{i_{NI}o} - s^{-*}_{i_{NI}o} = \sum_{j=1}^n \lambda_j^* x_{i_{NI}j} \geqslant 0$ for all $i_{NI}\in I^{NI}$, and $x_{i_Io} - \tilde{s}^{-*}_{i_Io} \geqslant \sum_{j=1}^n \lambda_j^* x_{i_Ij} \geqslant 0$ for all $i_I \in I^I$, implying that additive efficiency $\hat{\sigma}_o^*$ falls between zero and one. It can also be verified that $\hat{\sigma}_o^*$ is monotone decreasing in input and output slacks, and a larger value represents a better performance in reaching the efficient frontier. DMU$_o$ is called *additive efficient* if and only if $\hat{\sigma}_o^* = 1$, which also implies that all optimal slacks are zero. Otherwise it is defined as *additive inefficient*. It is easy to notice that if a DMU is *additive efficient*, then it will also be efficient under model (2).

3. Radial and additive integer-valued super-efficiency models

3.1. Radial super-efficiency model

For DMUs evaluated as efficient by additive model (3), they are also efficient if evaluated via radial model (2). We then turn to super-efficiency models to make a further discrimination. The idea of super-efficiency was first introduced by Andersen and Petersen (1993). We first consider standard radial super-efficiency models within an integer context. The resulting output-oriented super-efficiency for DMU$_o$ is assessed by the following VRS model (5).

$$
\begin{aligned}
\max \quad & \alpha_o^{VRS-Super} + \varepsilon\left(\sum_{i\in I} s^-_{io} + \sum_{r_{NI}\in O^{NI}} s^+_{r_{NI}o} + \sum_{r_I\in O^I}\left(s^+_{r_Io} + \tilde{s}^+_{r_Io}\right)\right) \\
s.t. \quad & x_{io} = \sum_{j=1, j\neq o}^n \lambda_j x_{ij} + s^-_{io}, \quad i \in I \\
& \alpha_o^{VRS-Super} y_{r_{NI}o} = \sum_{j=1, j\neq o}^n \lambda_j y_{r_{NI}j} - s^+_{r_{NI}o}, \quad r_{NI}\in O^{NI} \\
& \tilde{y}_{r_Io} = \sum_{j=1, j\neq o}^n \lambda_j y_{r_Ij} - s^+_{r_Io}, \quad r_I \in O^I \\
& \alpha_o^{VRS-Super} y_{r_Io} + \tilde{s}^+_{r_Io} = \tilde{y}_{r_Io}, \quad r_I \in O^I \\
& \tilde{y}_{r_Io} \in Z^+, \quad r_I \in O^I \\
& \sum_{j=1, j\neq o}^n \lambda_j = 1 \\
& \lambda_j \geqslant 0, \quad j = 1,\ldots,n, \quad j\neq o \\
& s^-_{io} \geqslant 0, \quad s^+_{r_{NI}o} \geqslant 0, \quad s^+_{r_Io} \geqslant 0, \quad \tilde{s}^+_{r_Io} \geqslant 0, \\
& i \in I, \quad r_{NI}\in O^{NI}, \quad r_I \in O^I
\end{aligned} \tag{5}
$$

As pointed out by some researchers (e.g., Seiford and Zhu, 1999; Chen, 2005), radial super-efficiency models under VRS with either orientation can be infeasible. Infeasibility is also possible to occur in model (5) when a DMU under evaluation cannot be projected onto the frontier formed by the rest DMUs. To combat this problem, we develop a new integer-valued super-efficiency model without the infeasibility issue.

3.2. Additive super-efficiency models

Suppose DMU$_o$ is additive efficient according to model (3). To obtain its super-efficiency score, we cannot simply modify additive model (3) by removing DMU$_o$ from the reference set. If we do that, the resulting model may not have a feasible solution (Du et al., 2010). Therefore, for an additive efficient DMU$_o$, we propose the following super-efficiency model

$$
\begin{aligned}
\hat{\alpha}_o^* = \min \quad & \hat{\alpha}_o = \sum_{i=1}^m t^-_{io} + \sum_{r=1}^s t^+_{ro} \\
s.t. \quad & \sum_{j=1, j\neq o}^n \lambda_j x_{ij} \leqslant x_{io} + t^-_{io}, i = 1,\ldots,m \\
& \sum_{j=1, j\neq o}^n \lambda_j y_{rj} \geqslant y_{ro} - t^+_{ro}, r = 1,\ldots,s \\
& t^-_{i_Io}, t^+_{r_Io} \in Z^+, i_I \in I^I, r_I \in O^I \\
& \sum_{j=1, j\neq o}^n \lambda_j = 1 \\
& \lambda_j, t^-_{i_{NI}o}, t^+_{r_{NI}o} \geqslant 0, \quad j = 1,\ldots,n,\ j\neq o, \\
& i_{NI}\in I^{NI}, \ r_{NI}\in O^{NI}
\end{aligned} \tag{6}
$$

After DMU$_o$ is removed from the reference set of model (3), we need to modify the constraints and objective of model (3) to get the corresponding super-efficiency model (6). The constraints should be modified because, in the super-efficiency model, we need to increase the inputs and decrease the outputs for DMU$_o$ to reach the frontier constructed by the remaining DMUs. We change the objective from maximization to minimization so that the resulting model is bounded.

We can also use a different objective function for model (6) so that the resulting super-efficiency model is unit-invariant, for example,

$$\hat{\beta}_o^* = \min \hat{\beta}_o = \frac{1}{m+s}\left(\sum_{i=1}^{m}\frac{t_{io}^-}{x_{io}} + \sum_{r=1}^{s}\frac{t_{ro}^+}{y_{ro}}\right) \tag{7}$$

The newly developed additive super-efficiency model (6) successfully avoids the infeasibility issue of radial super-efficiency model (5). This is proved as follows.

Theorem 1. *Additive VRS super-efficiency model (6) is always feasible.*

Proof. For any non-negative set of $\bar{\lambda}_j,\ j=1,\ldots,n,\ j\neq o$ satisfying $\sum_{j=1,j\neq o}^{n}\bar{\lambda}_j=1$, we define $\bar{t}_{i_{NI}o}^-=\max\left\{x_{i_{NI}o},\sum_{j=1,j\neq o}^{n}\bar{\lambda}_j x_{i_{NI}j}\right\}-x_{i_{NI}o}\geqslant 0$ for all $i_{NI}\in I^{NI}$, $\bar{t}_{r_{NI}o}^+=y_{r_{NI}o}-\min\left\{y_{r_{NI}o},\sum_{j=1,j\neq o}^{n}\bar{\lambda}_j y_{r_{NI}j}\right\}\geqslant 0$ for all $r_{NI}\in O^{NI}$, $\bar{t}_{i_Io}^-=\max\left\{x_{i_Io},\left[\sum_{j=1,j\neq o}^{n}\bar{\lambda}_j x_{i_Ij}\right]+1\right\}-x_{i_Io}\in Z^+$ for all $i_I\in I^I$, and $\bar{t}_{r_Io}^+=y_{r_Io}-\min\left\{y_{r_Io},\left[\sum_{j=1,j\neq o}^{n}\bar{\lambda}_j y_{r_Ij}\right]\right\}\in Z^+$ for all $r_I\in O^I$. Here the operation $[a]$ represents the greatest integer less than a. Then we obtain $x_{io}+\bar{t}_{io}^-\geqslant\sum_{j=1,j\neq o}^{n}\bar{\lambda}_j x_{ij}$ for all $i=1,\ldots,m$, and $y_{ro}-\bar{t}_{ro}^+\leqslant\sum_{j=1,j\neq o}^{n}\bar{\lambda}_j y_{rj}$ for all $r=1,\ldots,s$, indicating that $\{\lambda_j=\bar{\lambda}_j,\ j=1,\ldots,n,\ j\neq o;\ t_{io}^-=\bar{t}_{io}^-,\ i=1,\ldots,m;\ t_{ro}^+=\bar{t}_{ro}^+,\ r=1,\ldots,s\}$ is a feasible solution to model (6). Therefore, model (6) is always feasible. □

Let $\left\{\hat{\alpha}_o^*;\ \lambda_j^*,\ j=1,\ldots,n,\ j\neq o;\ t_{io}^{-*},\ i=1,\ldots,m;\ t_{ro}^{+*},\ r=1,\ldots,s\right\}$ be an optimal solution to model (6). Then following the definition on additive super-efficiency introduced in Du et al. (2010), we use $\hat{\delta}_o^*=\frac{\frac{1}{m}\sum_{i=1}^{m}(x_{io}+t_{io}^{-*})/x_{io}}{\frac{1}{s}\sum_{r=1}^{s}(y_{ro}-t_{ro}^{+*})/y_{ro}}\geqslant 1$ as the *additive super-efficiency* score for model (6). Note that $\hat{\delta}_o^*$ is monotone increasing in input/output slacks, indicating that a greater score represent a superior performance compared with other efficient units.

In both objective functions (4) and (7) and the definitions for additive efficiency and super-efficiency, $R_i^-=\max_j\{x_{ij}\}$ and $R_r^+=\max_j\{y_{rj}\}$ can be used instead of x_{io}, $i=1,\ldots,m$ and $y_{ro}, r=1,\ldots,s$, respectively. In that way, the positive requirement on all x_{io}, $i=1,\ldots,m$ and y_{ro}, $r=1,\ldots,s$ can be dropped.

4. Application to new business development

4.1. Data description

We use the data collected by the Inventor's Assistance Program (IAP) at the Canadian Innovation Centre (CIC) in Waterloo, Canada. The CIC is a non-profit agency that provides a variety of services to foster business development involving innovative inventions. Analysts in the IAP evaluate or judge the commercial quality of specific ideas and inventions submitted to the program by independent inventors before they have reached the market. The main purpose or contribution of this evaluation is to advise the potential entrepreneurs on whether and how to continue market-oriented efforts.

Analysts from IAP assess a specific product idea or invention before its debut on the market with the purpose of giving the potential entrepreneur advice, where possible for further improvement. Åstebro and Elhedhli (2006) mention that, to have a project evaluated, the entrepreneur fills out a questionnaire to introduce his/her background information, to briefly describe the new idea, and to provide supplementary documentation such as patent applications, sketches, and test reports. Avoiding personal contact with the entrepreneurs beyond the provided documentation, analysts from the IAP evaluated the ideas or inventions on 40 dimensions denoted by $v1$–$v40$. These 40 dimensions cover a wide range of *technical factors*, *production factors*, *market factors*, and *risk factors*. In practice, analysts compare a specific project with similar ones in their vast library of previous reviews, and then sort the projects into ordinal rankings: A for very good, B for average, C for exhibiting a critical flaw (Åstebro and Elhedhli, 2006).

The data for the present study were previously investigated in a number of studies. For example, Åstebro and Michela (2005) performed a cluster analysis and found that three variables significantly affect financial survival, namely, *anticipated stable demand*, *price required for profitability*, and *technical product maturity*. Furthermore, *the degree of competition* was also found to have a marginally significant influence. Later, Åstebro and Elhedhli (2006) applied the data set to test the ability of the heuristic to replicate the analysts' forecasts and to further examine how well this model predicts project outcomes. Their research results suggest that reasonably simple decision heuristics can perform well in a natural and rather difficult decision-making context.

Among the variables $v1$–$v40$ used by the IAP, several were dropped due to too many missing observations or changes in definitions over time, which disqualify them from use. In addition, to maintain the differentiating power of DEA models, we control the output measures within a reasonable number by only selecting those variables whose correlations with probability of commercialization are larger than 0.10, according to Spearman rank-order correlation results in Åstebro and Elhedhli (2006). Thus, the data used in the current paper involve 24 variables out of the original 40 evaluated by a three-point linguistic scale (A, B and C). Consider, for example, the early-stage Project 1. The analysts rate Project 1 on variable $v2$ (*functional performance*) as scale A (very good), corresponding to score 1. In other words, the analysts believe that Project 1 will effectively achieve the intended purpose. By excluding those projects with missing observations from the original data set, we finally obtain 490 projects with 436 failures and 54 successes spanning the period from 1989 to 1994. The definitions for all 24 variables used by the IAP, including a grouping of those variables, are displayed in Appendix A. In Appendix A, 12 of the 24 variables are labeled as *market factors*, and are further divided into subcategories of *demand* (described by 4 variables), *acceptability* (described by 4 variables), *competition* (described by 1 variable), and *effort* (described by 3 variables). The other 12 variables are labeled as *technical factors* (described by 3 variables), *production factors* (described by 3 variables), or *risk factors* (described by 6 variables). For a more detailed description on the factors and variables, refer to Åstebro and Gerchak (2001), Åstebro and Michela (2005), and Åstebro and Elhedhli (2006). These 24 variables reflect various angles analysts have taken to evaluate a specific innovation, and play an important role in determining business survival. Take the *trend of demand* for example. The greater the demand growth is, the greater the sales opportunities are, and thus the greater the chance for profitability is. Project 73 is a good example to demonstrate this. Analysts believe that the demand for Project 73 will be expected to rise in the lifetime of this idea, and give scale A (or score 1) on variable $v13$ (*trend of demand*). Later this project did in fact experience commercial success.

In previous studies on this early-stage business application, those 24 variables are given numerical values of −1, 0 or 1, corresponding to scale C, B or A. However, an invention or project could sometimes be evaluated as C+, B−, B+, or A−, together with C, B, A consisting a 7-point scale. Therefore, intermediate values for those assessments are assigned correspondingly, namely, −2/3 for C+, −1/3 for B−, 1/3 for B+, and 2/3 for A−. For a further and detailed explanation on the data description and collection, readers are referred to Åstebro and Michela (2005) and Simons and Åstebro (2010).

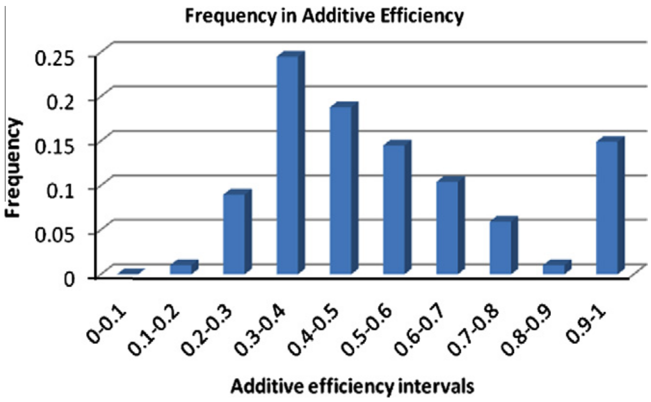

Fig. 1. Additive efficiency for all projects.

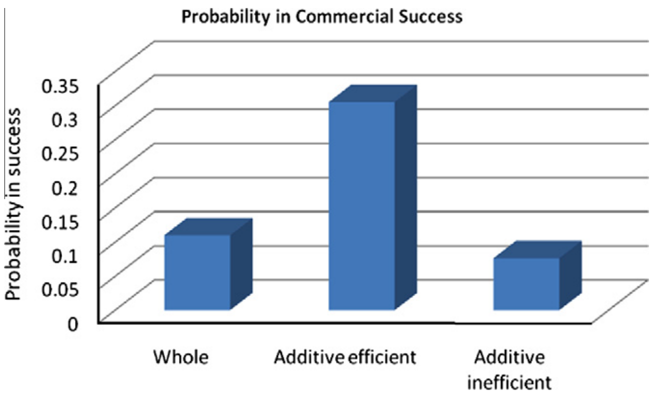

Fig. 2. Probability in commercial success.

4.2. Data processing

Although variables in previous research take values from −1, −2/3, −1/3, 0, 1/3, 2/3 and 1, they are chosen only to represent the afore-mentioned 7-point scale from C to A. We apply the integer DEA model in order to project the evaluated DMU's output measures to the original 7-point scale, which we think would make more practical sense than otherwise. Therefore it is reasonable to use integers from 1 to 7 in efficiency analysis of the current application instead of using the original values. Also, due to the data characteristics of the output variables (integers between 1 and 7), the performance targets for all 24 outputs should also be restricted to integers ranging from 1 to 7. Otherwise, if a fractional value is chosen for reference, it will have no corresponding position in the 7 scales, which can make the targets and the original data not comparable based on the same standard.

Note that for this data set, there is no input measure involved. To make the DEA method applicable, we assign a unified input with value 1 to all DMUs (or individual projects). Thus, each DMU has one input valued at 1 and 24 outputs with integer values between 1 and 7. Because only output variables are restricted to integers, it is appropriate to use the output-oriented models when taking a radial perspective. Moreover, non-unit-invariant models are chosen for analysis because all input/output measures in this application take values (integers ranging from 1 to 7) from the same magnitude.

4.3. Main results and discussion

We first analyze this data set using our additive integer model (3), and obtain additive efficiency and all 24 output slacks. Seventy-two projects are evaluated as *additive efficient*, which are also efficient according to radial model (2). All of the remaining 418 projects have efficiency less than 1, implying that they have non-zero slacks from additive model (3). The efficiency scores for all 490 observations have a diversified distribution, ranging from the lowest 0.18824 to the highest 1. Fig. 1 demonstrates the frequency of all projects falling into different efficiency intervals. We find that the efficiency interval between 0.3 and 0.4 has the highest frequency 0.245, which represents that approximately 24.5% of all projects, or 120 projects, have additive efficiency greater than 0.3 but less than or equal to 0.4. The efficiency intervals with the second and third highest frequency are (0.4, 0.5] (approximately 0.188, or 92 projects), and (0.9, 1] (approximately 0.149, or 73 projects), respectively. Also note that with the 72 efficient projects excluded, almost 80% of 418 *additive inefficient* projects have an efficiency score between 0.3 and 0.7.

Table 1
Radial super-efficiency.

Project	Sup-Radial	Sell	Project	Sup-Radial	Sell	Project	Sup-Radial	Sell
12	0.71429	0	47	0.85714	0	307	0.85714	0
33	0.71429	0	64	0.85714	0	321	0.85714	0
73	0.71429	1	82	0.85714	0	333	0.85714	0
99	0.71429	1	89	0.85714	0	336	0.85714	0
185	0.71429	1	105	0.85714	0	341	0.85714	0
211	0.71429	1	106	0.85714	0	348	0.85714	0
231	0.71429	1	109	0.85714	1	371	0.85714	1
259	0.71429	0	110	0.85714	0	378	0.85714	0
263	0.71429	0	134	0.85714	0	383	0.85714	0
285	0.71429	1	135	0.85714	0	387	0.85714	0
316	0.71429	0	163	0.85714	0	394	0.85714	1
325	0.71429	1	176	0.85714	0	409	0.85714	0
367	0.71429	0	177	0.85714	1	415	0.85714	1
438	0.71429	0	186	0.85714	0	421	0.85714	0
461	0.71429	0	189	0.85714	0	424	0.85714	1
157	0.75	0	190	0.85714	0	442	0.85714	1
220	0.75	1	202	0.85714	0	444	0.85714	0
289	0.75	0	206	0.85714	0	445	0.85714	1
293	0.75	0	222	0.85714	0	463	0.85714	1
452	0.75	1	228	0.85714	0	464	0.85714	0
3	0.85714	0	244	0.85714	1	466	0.85714	1
6	0.85714	1	249	0.85714	1	475	0.85714	0
36	0.85714	0	270	0.85714	0	476	0.85714	0
40	0.85714	0	275	0.85714	0	487	0.85714	0

Table 2
Additive super-efficiency.

Project	Sup-Additive	Sell	Project	Sup-Additive	Sell	Project	Sup-Additive	Sell
263	1.09804	0	461	1.03704	0	307	1.01818	0
211	1.07692	1	464	1.03704	0	321	1.01818	0
73	1.0752	1	475	1.03704	0	336	1.01818	0
99	1.07006	1	82	1.03067	0	341	1.01818	0
259	1.0566	0	134	1.03067	0	387	1.01818	0
285	1.0566	1	275	1.03067	0	409	1.01818	0
367	1.0566	0	6	1.02283	1	424	1.01818	1
438	1.0566	0	206	1.02283	0	442	1.01818	1
12	1.05	0	348	1.02283	0	444	1.01818	0
105	1.04186	0	33	1.01818	0	445	1.01818	1
383	1.04186	0	47	1.01818	0	463	1.01818	1
36	1.03704	0	106	1.01818	0	466	1.01818	1
89	1.03704	0	109	1.01818	1	476	1.01818	0
185	1.03704	1	110	1.01818	0	415	1.01664	1
220	1.03704	1	157	1.01818	0	3	1.01205	0
228	1.03704	0	163	1.01818	0	64	1.01205	0
231	1.03704	1	176	1.01818	0	135	1.01205	0
293	1.03704	0	177	1.01818	1	202	1.01205	0
316	1.03704	0	186	1.01818	0	270	1.01205	0
325	1.03704	1	189	1.01818	0	333	1.01205	0
371	1.03704	1	190	1.01818	0	421	1.01205	0
378	1.03704	0	244	1.01818	1	487	1.01205	0
394	1.03704	1	249	1.01818	1	40	1.00599	0
452	1.03704	1	289	1.01818	0	222	1.00599	0

According to original data set, 22 out of 72 efficient projects became a success. The chance of successful commercialization for all projects is 54/490, or approximately 11.0%, compared with the chance for efficient projects 22/72, or approximately 30.6%. Such an observation suggests that efficient projects have a much better chance (almost three times as much) in commercial success compared with the "all projects" pool. For the remaining 418 inefficient projects, only 32 of them thrive in the market, thus the chance is 32/418, or approximately 7.7%. This figure shows that probability in success for inefficient projects decreases greatly. Fig. 2 provides a clearer illustration and comparison of the probability in success for three groups, namely, the whole project pool, additive efficient and inefficient projects.

Next we focus on output slacks of the inefficient projects. Optimal slacks indicate those aspects that need to be improved to reach the efficient frontier. Based upon the total slacks in each output variable, the top five outputs in terms of "slack sum" are $v40$, $v38$, $v37$, $v33$, $v11$, which represent the most important five directions for performance improvement. We compare them with the top five variables on Spearman rank-order correlations with probability of commercialization, which are $v40$, $v38$, $v33$, $v23$ ($v39$), and$v37$ (Åstebro and Elhedhli, 2006). We further notice that the top five output variables contributing the most to efficiency assessment are very consistent with those correlating the most with probability of commercial success.

To further differentiate the 72 efficient projects, we first try radial super-efficiency model (4). Table 1 presents radial super-efficiency scores in column "Sup-Radial". The information on commercialization is demonstrated in column "Sell", with 1 representing success and 0 representing failure. It is seen from Table 1 that the radial super-efficiency has a fairly weak discrimination power. There are only three different super-efficiency values for all 72 efficient projects: 0.71429 for 15 projects, 0.75 for 5 projects, and 0.85714 for 52 projects.

Next we consider the additive super-efficiency model (6) to see if the same set of efficient DMUs can be better distinguished. Table 2 presents the related super-efficiency results for all efficient projects from high to low in column "Sup-Additive". The information on commercialization is listed in column "Sell". We observe that the differentiating power of additive super-efficiency is much stronger than that of radial super-efficiency. There are altogether fourteen different additive super-efficiency scores for all 72 efficient projects: 1.09804, 1.07692, 1.0752, 1.07006, 1.05, and 1.01664 respectively corresponds to one project; 1.0566 for 4 projects, 1.04186 for 2 projects, 1.03704 for 16 projects, 1.03067 for 3 projects, 1.02283 for 3 projects, 1.01818 for 28 projects, 1.01205 for 8 projects, and 1.00599 for 2 projects.

5. Conclusions

In many real-world DEA problems, some variables can only take integer values. Traditional DEA models assume that the input and output variables are continuous, and therefore the identified efficient benchmark targets are very likely to be fractional. To rectify this problem, the current paper proposes two types (radial and additive) of integer-valued efficiency and super-efficiency models.

We illustrate our models by an empirical efficiency evaluation of 490 potential business projects. We find that efficient projects have a much better chance in success than that of the "all projects" pool or inefficient projects. The results also demonstrate that additive super-efficiency models offer a higher discrimination power than radial super-efficiency models, where the latter does not consider slacks.

We note that while Tone (2002) develops the slacks-based super-efficiency measure, Ray (2008) uses the Nerlove–Luenberger (NL) measure of super-efficiency obtained from the directional distance function (DDF) (Chambers et al., 1996) in an empirical study on the airline industry. One possible study is to examine whether the DDF-based NL super-efficiency can be extended into integer-valued applications.

Finally, note that if we drop the constraint $\sum_{j=1}^{n}\lambda_j = 1$ in our models, we have models with constant returns to scale (CRS). Therefore, all the discussions can be applied to the CRS situation.

Acknowledgements

The authors are grateful for comments and suggestions by three anonymous reviewers, and thank the support by the National

Natural Science Foundation of China (Grant No. 71101108) and China Postdoctoral Science Foundation (Grant No. 20110490696).

Appendix A

Variable name	Description
Technical factors	
V1 Technical feasibility	Is the technical solution sound and complete?
V2 Functional performance	Will the innovation effectively achieve the intended purpose?
V3 Research and development	How great a burden is the remaining research and development required to bring the innovation to a marketable stage?
Production factors	
V8 Technology of production	Are the technology and skills required to produce the invention available?
V9 Tooling cost	How great a burden is the cost of production tooling required to meet the expected demand?
V10 Cost of production	Does production at a reasonable cost level appear possible?
Market factors demand	
V11 Need	Does the innovation solve a problem, ill a need or satisfy a want for the customer?
V13 Trend of demand	Will the demand for such an innovation be expected to rise, remain steady, or fall in the lifetime of this idea?
V15 Demand predictability	How closely will it be possible to predict sales?
V16 Product line potential	Can the innovation lead to other profitable products or services?
Acceptability	
V19 Compatibility	Is the innovation compatible with current attitudes and ways of doing things?
V22 Appearance	Does the appearance of the innovation convey a message of desirable qualities?
V23 Comparative functionality	Does this innovation work better than the alternatives?-or fulfill a unction not now provided?
V24 Durability	Will this innovation endure "long usage"?
Competition	
V26 Price	Does this innovation have a price advantage over its competitors?
Effort	
V29 Marketing research	How great an effort will be required to define the product and price that the final market will find acceptable?
V30 Promotion cost	Is the cost and effort of promotion to achieve market acceptance of the innovation in line with expected earnings?
V31 Distribution	How difficult will it be to develop or access distribution channels for the innovation?

Appendix A (*continued*)

Variable name	Description
Risk factors	
V33 Development risk	What degree of uncertainty is associated with complete successful development from the present condition of the innovation to the market ready state?
V35 Protection	Is it likely that worthwhile commercial protection will be obtainable for this innovation through patents, trade secrets or other means?
V37 Size of investment	Is the total investment required for the project likely to be obtainable?
V38 Potential sales	Is the sales volume for this particular innovation likely to be sufficient to justify initiating the project?
V39 Payback period	Will the initial investment be recovered in the early life of the innovation?
V40 Profitability	Will the expected revenue from the innovation provide more profits than other investment opportunities?

References

Åstebro, T., Elhedhli, S., 2006. The effectiveness of simple decision heuristics: forecasting commercial success for early-stage ventures. Management Science 52 (3), 395–409.

Åstebro, T., Gerchak, Y., 2001. Profitable advice: the value of information provided by Canada's Inventor's Assistance Program. Economics of Innovation and New Technology 10 (1), 45–72.

Åstebro, T., Michela, J.L., 2005. Predictors of the survival of innovations. The Journal of Product Innovation Management 22, 322–335.

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

Banker, R.J., Charnes, A., Cooper, W.W., 1984. Some models for estimating technical and scale inefficiencies in data envelopment analysis. Management Science 30 (9), 1078–1092.

Chambers, R.G., Chung, Y., Färe, R., 1996. Benefit and distance functions. Journal of Economic Theory 70 (2), 407–419.

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

Charnes, A., Cooper, W.W., Seiford, L.M., Stutz, J., 1982. A multiplicative model for efficiency analysis. Socio-Economic Planning Sciences 16 (5), 223–224.

Chen, Y., 2005. Measuring super-efficiency in DEA in the presence of infeasibility. European Journal of Operational Research 161, 545–551.

Chen, Y., Sherman, H.D., 2004. The benefits of non-radial vs. radial super-efficiency DEA: an application to burden-sharing amongst NATO member nations. Socio-Economic Planning Sciences 38 (4), 307–320.

Du, J., Liang, L., Zhu, J., 2010. A slacks-based measure of super-efficiency in data envelopment analysis: a comment. European Journal of Operational Research 204, 694–697.

Kazemi Matin, R., Kuosmanen, T., 2009. Theory of integer-valued data envelopment analysis under alternative returns to scale axioms. OMEGA 37, 988–995.

Kuosmanen, T., Kazemi Matin, R., 2009. Theory of integer-valued data envelopment analysis. European Journal of Operational Research 192, 658–667.

Lozano, S., Villa, G., 2006. Data envelopment analysis of integer-valued inputs and outputs. Computers and Operations Research 33 (10), 3004–3014.

Ray, S.C., 2008. The directional distance function and measurement of super-efficiency: an application to airlines data. Journal of the Operational Research Society 59, 788–797.

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

Simons, K.L., Åstebro, T., 2010. Entrepreneurs seeking gains: profit motives and risk aversion in inventors' commercialization decisions. Journal of Economics and Management Strategy 19 (4), 863–888.

Tone, K., 2001. A slacks-based measure of efficiency in data envelopment analysis. European Journal of Operational Research 130, 498–509.

Tone, K., 2002. A slacks-based measure of super-efficiency in data envelopment analysis. European Journal of Operational Research 143, 32–41.

Torgersen, A.M., Forsund, F.R., Kittelsen, S.A.C., 1996. Slack-adjusted efficiency measures and ranking of effcient units. Journal of Productivity Analysis 7, 379–398.