J Prod Anal
DOI 10.1007/s11123-014-0415-x

Composition versus decomposition in two-stage network DEA: a reverse approach

Dimitris K. Despotis · Gregory Koronakos · Dimitris Sotiros

© Springer Science+Business Media New York 2014

Abstract A two-stage production process assumes that the first stage transforms external inputs to a number of intermediate measures, which then are used as inputs to the second stage that produces the final outputs. The fundamental approaches to two-stage network data envelopment analysis are the multiplicative and the additive efficiency-decomposition approaches. Both they assume a series relationship between the two stages but they differ in the definition of the overall system efficiency as well as in the way they conceptualize the decomposition of the overall efficiency to the efficiencies of the individual stages. In this paper, we first show that the efficiency estimates obtained by the additive decomposition method are biased, by unduly favouring one stage against the other, while those obtained by the multiplicative method are not unique. Then, we present a novel approach to estimate unique and unbiased efficiency scores for the individual stages, which are then composed to obtain the efficiency of the overall system, by selecting the aggregation method a posteriori. Within the particularity of two-stage processes emerging from the conflicting role of the intermediate measures, we develop an envelopment model to locate the efficient frontier whose derivation from our primal multiplier efficiency assessment model is effectively justified. The results derived from our approach are compared with those obtained by the aforementioned basic methods on experimental data as well as on test data drawn from the literature. Similarities and dissimilarities in the results are rigorously justified.

D. K. Despotis (✉) · G. Koronakos · D. Sotiros
Department of Informatics, University of Piraeus, 80, Karaoli and Dimitriou, 18534 Piraeus, Greece
e-mail: despotis@unipi.gr

Keywords Data envelopment analysis (DEA) · Two-stage process · Network DEA · Decomposition · Composition · Efficient frontier

JEL Classification C61 · C67

1 Introduction

Data envelopment analysis (DEA) is a widely used technique for evaluating the performance of peer decision making units (DMUs) that use multiple inputs to produce multiple outputs. The two milestone DEA models, namely the CCR (Charnes et al. 1978) and the BCC (Banker et al. 1984) models, have become standards in the literature of performance measurement under the assumptions of constant and variable returns-to-scale respectively. Conventional DEA models treat the DMUs as single stage production processes that transform some external inputs to final outputs. In such a setting, the internal structure of the DMUs is not taken into consideration. However, a significant number of studies has focused on assessing efficiency in multi-stage production processes, where outputs from some stages, characterized as intermediate products, are used either as inputs to the other stages or as external outputs of the production process. Färe and Grosskopf (1996) were among the first to deal with efficiency assessments in such processes, represented as network activity analysis models. Castelli et al. (2010) provide a comprehensive categorized overview of models and methods developed for different multi-stage production architectures. In this paper, however, we focus on the typical architecture of a two-stage production process, which assumes that the external inputs entering the first stage of the process are

Published online: 16 September 2014

transformed to a number of intermediate measures that are then used as inputs to the second stage to produce the final outputs. In this architecture, nothing but the external inputs to the first stage enters the system and nothing but the outputs of the second stage leaves the system. Kao and Hwang (2008) introduced an innovative approach by taking into account a series relationship of the two stages and developed a model to estimate the overall efficiency of the production process as the product of the efficiencies of the two individual stages. Their approach is based on the reasonable assumption that the values of the intermediate measures (virtual intermediate measures) are the same, no matter if they are considered as outputs of the first stage or inputs to the second stage. As they noted, the decomposition of the overall efficiency to the stage efficiencies may be not unique. In order to check the uniqueness, they proposed a post-optimality procedure, to obtain the largest first (or second) stage efficiency score while keeping the overall efficiency unchanged. Liang et al. (2008) view the efficiency assessments in two-stage process in terms of a game approach. Maintaining the series relationship between the two stages, Chen et al. (2009) introduced the additive efficiency decomposition in two-stage processes. They derive the overall efficiency of the production process as a weighted average of the efficiencies of the individual stages. Their modeling approach facilitates the linearization of a non-linear mathematical program by assuming that the weights of the two stages derive endogenously by the optimization process. As it will be made clear in the following section, this assumption leads to biased efficiency assessments. An issue investigated further in the literature is the derivation of the efficient frontier in two-stage DEA. Chen et al. (2010) pointed out that adjusting the inputs and the outputs by the efficiency scores is not sufficient to yield a frontier projection, when the additive decomposition model is assumed. They developed instead a model for deriving the efficient frontier within the Kao and Hwang (2008) multiplicative framework. The inability of the two-stage DEA models to locate correctly the efficient frontier, as it is the case with standard DEA, is further examined in Chen et al. (2013). There, it was demonstrated that under general network structures, the multiplier and the envelopment network DEA models are two different approaches, thus, alternative methods to overcome this deficiency were reviewed.

In this paper we revisit the additive and the multiplicative efficiency-decomposition methods to discuss the aforementioned shortcomings. Then, based on a reverse perspective on how to obtain and aggregate the stage efficiencies, that of the *composition* as opposed to the *decomposition*, we develop a novel approach to two-stage network DEA that overcomes the deficiencies of the decomposition methods. Selecting an output orientation for the first stage and an input orientation for the second stage, we show that it is possible to obtain unbiased efficiency scores for the two stages in a bi-objective optimization framework. We propose two alternative models by employing different scalarizing functions in a multiobjective linear programming (MOLP) model. Firstly, we aggregate additively the two objectives in a single objective that locates an extreme (vertex) Pareto-optimal solution. Then, we develop a min–max model that provides unique efficiency scores by locating a point on the Pareto front, not necessarily extreme. In the latter case, the stage efficiencies obtained are more balanced. The individual efficiency scores are then used to calculate the overall efficiency of the production process, by selecting the aggregation (composition) method a posteriori. As the conflicting role of the intermediate measures gives a peculiar character to two-stage processes that obscures the standard DEA premises, we develop an envelopment model to locate the efficient frontier, whose derivation from our primal multiplier model is justified.

The paper unfolds as follows. In section two, we provide a critical review of the additive and the multiplicative efficiency decomposition approaches and we discuss their inherent limitations and shortcomings. In section three, we introduce a novel approach to assess the individual efficiencies of the two stages and the overall efficiency of the two-stage process, which effectively overcomes the shortcomings of the additive (Chen et al. 2009) and the multiplicative (Kao and Hwang 2008) decomposition methods. In section four, we provide extensive comparisons of our approach with the aforementioned two decomposition methods, on the basis of experimental data as well as with test data drawn from the literature. We provide also rigorous justifications for the similarities and the differentiations observed in the results. In section five, we introduce an envelopment model to derive the efficient frontier in two-stage DEA. It is linked to—and developed on the basis of—our primal multiplier efficiency assessment model. Conclusions are drawn in section six.

2 The decomposition approach to two-stage DEA: a critical review

Consider the generic case where each DMU transforms some external inputs X to final outputs Y via the intermediate measures Z with a two-stage process, as depicted in Fig. 1.

Assume n DMUs ($j = 1,\ldots,n$), each using m external inputs x_{ij}, $i = 1,\ldots,m$ in the first stage to produce q outputs z_{pj}, $p = 1,\ldots,q$ from that stage. The outputs obtained from

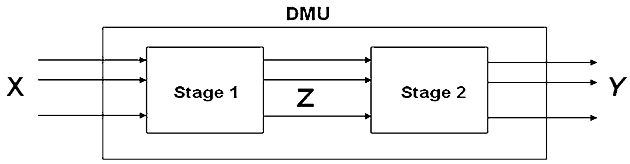

Fig. 1 The architecture of a generic two-stage process

the first stage are then used as inputs to the second stage to produce s final outputs y_{rj}, $r = 1,\ldots,s$. In this basic setting, nothing but the external inputs to the first stage enters the system and nothing but the outputs of the second stage leaves the system. Let us introduce the following basic notation:

$j \in J = \{1, \ldots, n\}$: The index set of the n DMUs.
$j_0 \in J$: Denotes the evaluated DMU.
$X_j = (x_{ij}, i = 1, \ldots, m)$: The vector of external inputs used by DMU_j, $j \in J$
$Z_j = (z_{pj}, p = 1, \ldots, q)$: The vector of intermediate measures for DMU_j, $j \in J$
$Y_j = (y_{rj}, r = 1, \ldots, s)$: The vector of final outputs produced by DMU_j, $j \in J$
$\eta = (\eta_1, \ldots, \eta_m)$: The vector of weights for the external inputs in the fractional model.
$v = (v_1, \ldots, v_m)$: The vector of weights for the external inputs in the linear model.
$\varphi = (\varphi_1, \ldots, \varphi_q)$: The vector of weights for the intermediate measures in the fractional model.
$w = (w_1, \ldots, w_q)$: The vector of weights for the intermediate measures in the linear model.
$\omega = (\omega_1, \ldots, \omega_s)$: The vector of weights for the final outputs in the fractional model.
$u = (u_1, \ldots, u_s)$: The vector of weights for the final outputs in the linear model.
e_j^o: The overall efficiency of DMU_j, $j \in J$
e_j^1: The efficiency of the first stage for DMU_j, $j \in J$
e_j^2: The efficiency of the second stage for DMU_j, $j \in J$

A major characteristic of the decomposition approach is that, apart from the definition of the efficiency of the two individual stages (stage efficiencies), it premises the definition of the overall efficiency of the DMU together with a model to decompose the overall efficiency to the stage efficiencies. Then, the efficiency scores of the two stages derive as offspring of the overall efficiency of the unit. The two basic decomposition methods dominating the literature on two-stage DEA, i.e. the multiplicative method of Kao and Hwang (2008) and the additive method of Chen et al. (2009) assume the same definitions of stage efficiencies but they differ substantially in the definition of the overall efficiency as well as in the decomposition model used.

Consider the basic input oriented CRS-DEA models that estimate the stage-1, the stage-2 and the overall efficiency for the evaluated unit j_0 independently:

Stage 1:

$$\begin{aligned} &\max \frac{\varphi Z_{j_0}}{\eta X_{j_0}} \\ &s.t. \\ &\frac{\varphi Z_j}{\eta X_j} \le 1, \quad j = 1, \ldots, n \\ &\eta \ge 0, \varphi \ge 0 \end{aligned} \tag{1}$$

Stage 2:

$$\begin{aligned} &\max \frac{\omega Y_{j_0}}{\hat{\varphi} Z_{j_0}} \\ &s.t. \\ &\frac{\omega Y_j}{\hat{\varphi} Z_j} \le 1, \quad j = 1, \ldots, n \\ &\hat{\varphi} \ge 0, \omega \ge 0 \end{aligned} \tag{2}$$

Overall:

$$\begin{aligned} &\max \frac{\omega Y_{j_0}}{\eta X_{j_0}} \\ &s.t. \\ &\frac{\omega Y_j}{\eta X_j} \le 1, \quad j = 1, \ldots, n \\ &\eta \ge 0, \omega \ge 0 \end{aligned} \tag{3}$$

In order to link the efficiency assessments of the two stages, it is universally accepted that the weights associated with the intermediate measures are the same (i.e. $\hat{\varphi} = \varphi$), no matter if these measures are considered as outputs of the first stage or inputs to the second stage.

2.1 The multiplicative method

In the multiplicative method introduced by Kao and Hwang (2008), the overall efficiency and the stage efficiencies of the DMU_j are defined as follows:

$$e_j^o = \frac{\omega Y_j}{\eta X_j}, e_j^1 = \frac{\varphi Z_j}{\eta X_j}, e_j^2 = \frac{\omega Y_j}{\varphi Z_j} \tag{4}$$

whereas the decomposition model used is

$$e_j^o = \frac{\omega Y_j}{\eta X_j} = \frac{\varphi Z_j}{\eta X_j} \cdot \frac{\omega Y_j}{\varphi Z_j} = e_j^1 \cdot e_j^2 \tag{5}$$

i.e. the overall efficiency is the *square geometric average* of the stage efficiencies.

Given the above definitions, the model below assesses the overall efficiency of the evaluated unit j_0:

$$e_{j0}^{o} = \max \frac{\omega Y_{j_0}}{\eta X_{j_0}}$$
$$s.t.$$
$$\frac{\varphi Z_j}{\eta X_j} \leq 1, \quad j = 1, \ldots, n \tag{6}$$
$$\frac{\omega Y_j}{\varphi Z_j} \leq 1, \quad j = 1, \ldots, n$$
$$\varphi \geq 0, \eta \geq 0, \omega \geq 0$$

Notice that the constraints $\omega Y_j / \eta X_j \leq 1, j = 1, \ldots, n$ are redundant and, thus, omitted. Model (6) is a fractional linear program that can be modeled and solved as a linear program by applying the Charnes and Cooper (1962) transformation (C–C transformation hereafter), as follows:

$$e_{j0}^{o} = \max uY_{j_0}$$
$$s.t.$$
$$vX_{j_0} = 1$$
$$wZ_j - vX_j \leq 0, \quad j = 1, \ldots, n \tag{7}$$
$$uY_j - wZ_j \leq 0, \quad j = 1, \ldots, n$$
$$v \geq 0, w \geq 0, u \geq 0$$

Once an optimal solution (v^*, w^*, u^*) of model (7) is obtained, the overall efficiency and the stage efficiencies are calculated as follows:

$$e_{j_o}^{o} = u^* Y_{j_0}, \quad e_{j_o}^{1} = w^* Z_{j_0}, \quad e_{j_o}^{2} = \frac{e_{j_o}^{o}}{e_{j_o}^{1}} \tag{8}$$

Notice that the overall efficiency is obtained as the optimal value of the objective function, the stage-1 efficiency is given by the virtual intermediate measure, whereas the stage-2 efficiency derives as offspring of the overall and stage-1 efficiencies. A major shortcoming of the multiplicative method is that the decomposition of the overall efficiency to the stage efficiencies is not unique. Indeed, as the term wZ_{j_0} does not appear in neither the objective function or in the normalization constraint, its value may vary and still maintain the optimal value of the objective function (i.e. the overall efficiency) and the inequality constraints of model (7). That is why Kao and Hwang (2008) propose solving a pair of linear programs, in a post-optimality phase, to obtain extreme values for $e_{j_o}^{1}$ and $e_{j_o}^{2}$ while maintaining the overall efficiency score obtained by model (7). The argument is that one might wish giving priority to the first or the second stage in the efficiency assessments. Although there is a rationale in this argument, the non-uniqueness of the decomposition is still a problem, especially in the case that no priority is conceived by the management. A number of examples verified the non-uniqueness of the multiplicative decomposition.

2.2 The additive method

In the additive decomposition method introduced by Chen et al. (2009), the overall efficiency and the stage efficiencies of the DMUj are defined as follows:

$$e_j^o = \frac{\omega Y_j + \varphi Z_j}{\eta X_j + \varphi Z_j}, e_j^1 = \frac{\varphi Z_j}{\eta X_j}, e_j^2 = \frac{\omega Y_j}{\varphi Z_j} \tag{9}$$

The definition of the stage efficiencies are the same as in the multiplicative method, but the additive method differentiates in the definition of the overall efficiency. Notably, in (9) the intermediate measures appear in both terms of the fraction that defines the overall efficiency, meaning that they are considered as inputs and as outputs simultaneously. The decomposition model used is as follows:

$$\frac{\omega Y_j + \varphi Z_j}{\eta X_j + \varphi Z_j} = t_j^1 \frac{\varphi Z_j}{\eta X_j} + t_j^2 \frac{\omega Y_j}{\varphi Z_j}, \quad t_j^1 + t_j^2 = 1 \tag{10}$$

i.e. the overall efficiency is *a weighted arithmetic average* of the stage efficiencies. The functional forms of the weights derive by solving the system (10) for t_j^1 and t_j^2, as follows:

$$t_j^1 = \frac{\eta X_j}{\eta X_j + \varphi Z_j}, \quad t_j^2 = \frac{\varphi Z_j}{\eta X_j + \varphi Z_j} \tag{11}$$

Notably, as the weights are functions of the virtual measures, they depend on the unit being evaluated and, obviously, they generally differentiate from one unit to another.

Given the above definitions, the model below assesses the overall efficiency of the evaluated unit j_0:

$$e_{j0}^{o} = \max \frac{\omega Y_{j_0} + \varphi Z_{j_0}}{\eta X_{j_0} + \varphi Z_{j_0}}$$
$$s.t.$$
$$\frac{\varphi Z_j}{\eta X_j} \leq 1, \quad j = 1, \ldots, n \tag{12}$$
$$\frac{\omega Y_j}{\varphi Z_j} \leq 1, \quad j = 1, \ldots, n$$
$$\eta \geq 0, \varphi \geq 0, \omega \geq 0$$

Applying the C–C transformation to the linear fractional model (12), the following linear program is modeled and solved:

$$e_{j0}^{o} = \max uY_{j_0} + wZ_{j_0}$$
$$s.t.$$
$$vX_{j_0} + wZ_{j_0} = 1$$
$$wZ_j - vX_j \leq 0, \quad j = 1, \ldots, n \tag{13}$$
$$uY_j - wZ_j \leq 0, \quad j = 1, \ldots, n$$
$$v \geq 0, w \geq 0, u \geq 0$$

Once an optimal solution (v^*, w^*, u^*) of model (13) is obtained, the overall efficiency and the stage efficiencies are calculated as follows:

$$
\begin{aligned}
e^o_{j_o} &= u^* Y_{j_0} + w^* Z_{j_0} \\
t^1_{j_0} &= v^* X_{j_0}, \quad t^2_{j_0} = w^* Z_{j_0} \\
e^1_{j_o} &= \frac{w^* Z_{j_0}}{v^* X_{j_0}} = \frac{t^2_{j_0}}{t^1_{j_0}} \\
e^2_{j_o} &= \frac{e^o_{j_o} - t^1_{j_0} e^1_{j_o}}{t^2_{j_0}} = \frac{u^* Y_{j_0}}{w^* Z_{j_0}}
\end{aligned}
\tag{14}
$$

The overall efficiency $e^o_{j_0}$ is obtained as the optimal value of the objective function, the weight $t^1_{j_0}$ is obtained as the optimal virtual input, the weight $t^2_{j_0}$ is obtained as the optimal virtual intermediate measure, the efficiency of the first stage $e^1_{j_0}$ is given by the ratio of the two weights, whereas the efficiency of the second stage $e^2_{j_0}$ is obtained as offspring of $e^o_{j_0}, e^1_{j_0}, t^1_{j_0}, t^2_{j_0}$.

The argument given in Chen et al. (2009) for the weights t^1_j and $t^2_j$is that they represent the relative contribution of the two stages to the overall performance of the DMU. The "size" of each stage, as measured by the portion of total resources devoted to each stage, is assumed to reflect their relative contribution to the overall efficiency of the DMU. However, as long as the weights derive from the optimal solution of (13), they depend on the DMU being evaluated and, generally, they are different for different DMUs. Thus, the "size" of a stage is not an objective reality, as it is viewed differently from each DMU. But this is not the only peculiarity emerging from the definition of the weights. Indeed, from the definition of the weights (11), as well as (14) holds that

$$\frac{t^2_j}{t^1_j} = \frac{w Z_j}{v X_j} = e^1_j \leq 1$$

i.e. $t^2_j \leq t^1_j$. This is a shortcoming of the additive decomposition model (13), as it biases the efficiency assessments in favor of the second stage. Indeed, the maximum value that t^2_j can attain is 0.5 and e^2_j increases (e^1_j decreases) as t^2_j decreases. As long as the individual efficiency scores are biased, the overall efficiency score is biased as well.

3 The composition approach: a reverse perspective

In this section we introduce a bias-free approach to assess unique efficiency scores for the two stages, which are then aggregated to get the overall efficiency score of the evaluated unit. Unlike the decomposition methods presented in the previous section, our method does not require an a priori definition of the overall efficiency. This grants our approach the flexibility to select the aggregation method a posteriori. Let us call this approach "*the composition approach*" as opposed to the decomposition approach. Similarly to the other methods, we define the efficiency of the two stages as follows:

$$\hat{e}^1_j = \frac{\varphi Z_j}{\eta X_j}, \hat{e}^2_j = \frac{\omega Y_j}{\varphi Z_j}$$

3.1 Constant returns to scale

Consider the reciprocal of model (1), that is the output-oriented CRS-DEA model for the first-stage and the input-oriented CRS-DEA model (2) for the second-stage, where the same intermediate weights are assumed for both stages:

Stage I: Output-oriented

$$
\begin{aligned}
&\min \frac{\eta X_{j_0}}{\varphi Z_{j_0}} \\
&s.t. \\
&\frac{\eta X_j}{\varphi Z_j} \geq 1, \quad j = 1, \ldots, n \\
&\eta \geq 0, \varphi \geq 0
\end{aligned}
\tag{15}
$$

Stage II: Input-oriented

$$
\begin{aligned}
&\max \frac{\omega Y_{j_0}}{\varphi Z_{j_0}} \\
&s.t. \\
&\frac{\omega Y_j}{\varphi Z_j} \leq 1, \quad j = 1, \ldots, n \\
&\varphi \geq 0, \omega \geq 0
\end{aligned}
\tag{16}
$$

As mentioned earlier, models (15) and (16) provide the independent efficiency scores $1 \big/ E^1_{j_0}, E^2_{j_0}$ for the first and the second stage respectively. Appending the constraints of model (15) to model (16) and vice versa we get the following augmented models (17) and (18) for the first and the second stage respectively:

Stage I: Output-oriented

$$
\begin{aligned}
&\min \frac{\eta X_{j_0}}{\varphi Z_{j_0}} \\
&s.t. \\
&\frac{\eta X_j}{\varphi Z_j} \geq 1, \quad j = 1, \ldots, n \\
&\frac{\omega Y_j}{\varphi Z_j} \leq 1, \quad j = 1, \ldots, n \\
&\eta \geq 0, \varphi \geq 0, \omega \geq 0
\end{aligned}
\tag{17}
$$

Stage II: Input-oriented

$$\begin{aligned}
&\max \frac{\omega Y_{j_0}}{\varphi Z_{j_0}}\\
&s.t.\\
&\frac{\eta X_j}{\varphi Z_j} \geq 1, \quad j = 1,\ldots,n\\
&\frac{\omega Y_j}{\varphi Z_j} \leq 1, \quad j = 1,\ldots,n\\
&\eta \geq 0, \varphi \geq 0, \omega \geq 0
\end{aligned} \tag{18}$$

Notice that an optimal solution of model (15) is also optimal in model (17). Indeed, one can always choose small enough values for ω in model (17) to make any optimal solution of model (15) feasible, yet optimal, in model (17). Analogously, an optimal solution of model (16) is also optimal in model (18), as one can choose large enough values for η in model (18) to make any optimal solution of model (16) feasible, yet optimal, in model (18).

Models (17) and (18) have common constraints and, thus, can be jointly considered as a bi-objective program:

$$\begin{aligned}
&\min \frac{\eta X_{j_0}}{\varphi Z_{j_0}}\\
&\max \frac{\omega Y_{j_0}}{\varphi Z_{j_0}}\\
&s.t.\\
&\frac{\eta X_j}{\varphi Z_j} \geq 1, \quad j = 1,\ldots,n\\
&\frac{\omega Y_j}{\varphi Z_j} \leq 1, \quad j = 1,\ldots,n\\
&\eta \geq 0, \varphi \geq 0, \omega \geq 0
\end{aligned} \tag{19}$$

Applying the C–C transformation, model (19) can be formulated and solved as a multiobjective linear program (MOLP). The correspondence of variables is: $v = t\eta$, $u = t\omega$, $w = t\varphi$ where t is a scalar variable such that: $t\varphi Z_{j_0} = 1$.

$$\begin{aligned}
&E^1_{j_0} = \min vX_{j_0}\\
&E^2_{j_0} = \max uY_{j_0}\\
&s.t.\\
&wZ_{j_0} = 1\\
&wZ_j - vX_j \leq 0, \quad j = 1,\ldots,n\\
&uY_j - wZ_j \leq 0, \quad j = 1,\ldots,n\\
&v \geq 0, w \geq 0, u \geq 0
\end{aligned} \tag{20}$$

Optimizing the first and the second objective function separately one gets the independent efficiency scores of the two stages $(1\big/E^1_{j_0} \leq 1, E^2_{j_0} \leq 1)$. In terms of MOLP, the vector$(E^1_{j_0} \geq 1, E^2_{j_0} \leq 1)$ constitutes the ideal point of the bi-objective program (20) in the objective functions space. Thus, the efficiencies of the two stages can be obtained by solving the MOLP (20). However, as the ideal point is not generally attainable, solving a MOLP means finding non-dominated feasible solutions in the variable space that are mapped on the Pareto front in the objective functions space, i.e. solutions that they cannot be altered to increase the value of one objective function without decreasing the value of at least one other objective function. Among the different approaches to solving a MOLP, we adopt the a priori preference aggregation approach, which can readily host preference-free as well as preference intensive assessments. In the following, we develop our models for preference-free (neutral) efficiency assessments, i.e. without assuming any preference from the analyst giving priority to one of the two stages. Incorporation of prioritizing preferences is straightforward. A usual approach in solving a MOLP is the scalarizing approach, which transforms the MOLP in a single-objective LP, whose optimal solution is a Pareto optimal (non-dominated) solution of the MOLP. Aggregating additively the objective functions and introducing a distance function are two alternative methods to build the scalarizing function. We present both cases in the following, as they possess different properties.

Aggregating the two objective functions additively, we get the following single-objective program:

$$\begin{aligned}
&\min vX_{j_0} - uY_{j_0}\\
&s.t.\\
&wZ_{j_0} = 1\\
&wZ_j - vX_j \leq 0, \quad j = 1,\ldots,n\\
&uY_j - wZ_j \leq 0, \quad j = 1,\ldots,n\\
&v \geq 0, w \geq 0, u \geq 0
\end{aligned} \tag{21}$$

Once an optimal solution (v^*, w^*, u^*) of model (21) is obtained, the efficiency scores for unit j_c in the first and the second stage are respectively:

$$\hat{e}^1_{j_0} = \frac{w^* Z_{j_0}}{v^* X_{j_0}} = \frac{1}{v^* X_{j_0}}, \quad \hat{e}^2_{j_0} = \frac{u^* Y_{j_0}}{w^* Z_{j_0}} = u^* Y_{j_0} \tag{22}$$

The optimal value of the objective function in (21) is $v^* X_{j_0} - u^* Y_{j_0} \geq 0$. The unit j_0 is efficient in both stages and, thus, overall efficient, if and only if the optimal value of the objective function is zero. Otherwise it is overall inefficient. Indeed, if $v^* X_{j_0} - u^* Y_{j_0} = 0$ then, as $w^* Z_{j_0} = 1$ and $uY_j \leq wZ_j \leq vX_j$ for every j, it holds that $v^* X_{j_0} = w^* Z_{j_0} = u^* Y_{j_0} = 1$, i.e. $\hat{e}^1_{j_0} = 1$, $\hat{e}^2_{j_0} = 1$. Model (21) does not provide a direct measure of the overall efficiency, as it is the case in the multiplicative model (7) and the additive model (13), but it does discriminate among overall efficient and inefficient units, a property that is closely related to the standard additive DEA model. However, it is the normalization constraint $wZ_{j_0} = 1$, on the intermediate measures in (21), that allows us to infer on the efficiency scores of the individual stages, as given in (22). This is the key that enables us to assess the efficiencies of the two stages simultaneously without the need to assume weights for the two stages. Hence, our approach is “neutral”, as opposed to the Chen’s et al. (2009) one, where the endogenous weights assumed for the individual stages favor the second stage against the first one.

The optimal solution (v^*, w^*, u^*) of model (21) is a Pareto-optimal solution of the MOLP (20) whereas the optimal vector $(v^*X_{j_0}, u^*Y_{j_0})$ is a non-dominated point on the Pareto front in the objective functions space of (20). This is a direct implication of the Geoffrion’s (1968) theorem, which states that: given a multi-objective LP model $\{\min g_j(x), \quad j = 1,\ldots,n / x \in X, x \geq 0\}, x^*$ is a Pareto-optimal (efficient) solution for this model if and only if there are $\{t_j > 0, \quad j = 1,\ldots,n / \sum_{j=1}^{n} t_j = 1)$ such that $x^*$is optimal for the scalar LP model $\{\min \sum_{j=1}^{n} t_j g_j\ (x) / x \in X,\ x \geq 0\}$. Getting advantage of this property, one can scan the Pareto front and get alternative Pareto-optimal solutions by solving the following model for different values of the parameter t with $0 < t < 1$:

$$
\begin{aligned}
& \min\ tvX_{j_0} - (1-t)uY_{j_0} \\
& s.t. \\
& wZ_{j_0} = 1 \\
& wZ_j - vX_j \leq 0, \quad j = 1,\ldots,n \\
& uY_j - wZ_j \leq 0, \quad j = 1,\ldots,n \\
& v \geq 0, w \geq 0, u \geq 0
\end{aligned}
\tag{23}
$$

Notice that model (23) provides only extreme points on the Pareto front. Notice also that the same Pareto-optimal point can be obtained for a range of values of t, the so called indifference range. Thus, the solution obtained from model (21) by way of its unweighted scalar objective function can be obtained as well by giving different priorities (weights) to the two terms of the objective function within their indifference range. Figure 2 below is a general representation of the objective functions space of the MOLP (20) for an evaluated unit (X_0, Z_0, Y_0). Actually, it is the plane in the three-dimensional space (vX, wZ, uY) that is vertical to the axis wZ at $wZ_0 = 1$. The point (E^1, E^2)represents the ideal point, whereas the points A, B, C and D are the alternative Pareto optimal extreme points derived by the parametric model (23) for different values of the parameter t. The crooked line ABCD represents the Pareto front in the objective functions space. The dotted line passing from the point B has slope 1 and depicts the objective function of model (21), which, when minimized for the optimal solution (v^*, w^*, u^*), takes the non-negative value $v^*X_0 - u^*Y_0 = b > 0$ and locates the point B on the Pareto front.

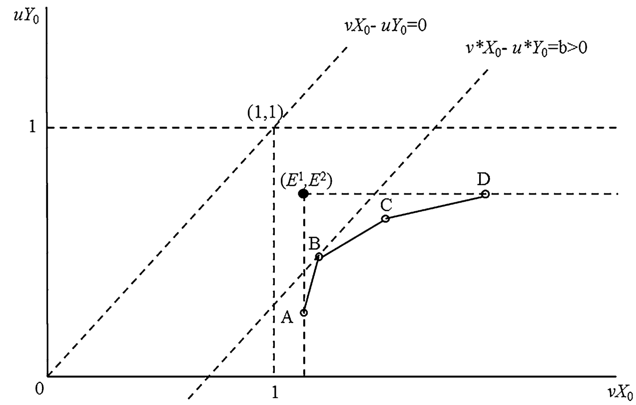

Fig. 2 The Pareto front of MOLP (20) and the optimal solution of (21)

Although it is not very likely to occur in practice, the Pareto optimal point derived by model (21) and, thus, the efficiency scores of the two stages might be non-unique. This is the case where a segment of the Pareto front has slope 1, i.e. when it is parallel to the objective function line. For example, if the segment BC defined by the two successive Pareto optimal points B and C was parallel to the objective function line, then B, C and any convex combination of them would be optimal in terms of model (21). The uniqueness of the Pareto-optimal point $(v^*X_{j_0}, u^*Y_{j_0})$ and, thus, the uniqueness of the optimal efficiency scores of the two stages derived by model (21), can be tested by minimizing vX_0 and maximizing uY_0 subject to the constraints of (21) plus the constraint $vX_{j_0} - uY_{j_0} \leq v^*X_{j_0} - u^*Y_{j_0}$.

Model (21) is equivalent to finding an optimal solution that locates a point on the Pareto front at a minimum sum of the deviations $vX_{j_0} - 1$ and $1 - u\Upsilon_{j_0}$ (L_1 norm) of $(vX_{j_0}, u\Upsilon_{j_0})$ from the boundary point (1,1) in the objective function space. Next, we employ the unweighted Tchebycheff norm (L_∞ norm) to locate a unique solution on the Pareto front by minimizing the maximum of the deviations $vX_{j_0} - E^1_{j_0}$ and $E^2_{j_0} - u\Upsilon_{j_0}$ of $(vX_{j_0}, u\Upsilon_{j_0})$ from the ideal point $\left(E^1_{j_0}, E^2_{j_0}\right)$. This is accomplished by the following minmax model, where δ denotes the largest deviation:

$$
\begin{aligned}
& \min\ \delta \\
& s.t. \\
& vX_{j_0} - \delta \leq E^1_{j_0} \\
& u\Upsilon_{j_0} + \delta \geq E^2_{j_0} \\
& wZ_{j_0} = 1 \\
& wZ_j - vX_j \leq 0, \quad j = 1,\ldots,n \\
& u\Upsilon_j - wZ_j \leq 0, \quad j = 1,\ldots,n \\
& v \geq 0, w \geq 0,\ u \geq 0, \delta \geq 0
\end{aligned}
\tag{24}
$$

Solving model (24) means searching for a solution where the deviations from the ideal point are equal and minimized. As depicted in Fig. 3, the minmax solution is D, being the intersection of the Pareto front and a ray from the ideal point (E^1, E^2) with slope (-1). The main advantage of model (24) over model (21) and the decomposition models (7) and (13) is that it provides a unique point, not necessarily extreme (vertex), on the Pareto front, i.e. unique efficiency scores for the two stages. Once an optimal solution (v^*, w^*, u^*) of model (24) is obtained, the stage efficiency scores for unit j_0 are as in (22).

Considering the weighted Tchebycheff distance, the following parametric minmax model searches for a solution where the weighted deviations $t(vX_{j_0} - E^1_{j_0})$ and $(1-t)(E^2_{j_0} - u\Upsilon_{j_0})$, with $0 < t < 1$, are equal and minimized.

$$\begin{aligned}
&\min\ \delta\\
&s.t.\\
&tvX_{j_0} - \delta \le tE^1_{j_0}\\
&(1-t)u\Upsilon_{j_0} + \delta \ge (1-t)E^2_{j_0}\\
&wZ_{j_0} = 1\\
&wZ_j - vX_j \le 0, \quad j = 1,\ldots,n\\
&u\Upsilon_j - wZ_j \le 0, \quad j = 1,\ldots,n\\
&v \ge 0, w \ge 0,\ u \ge 0, \delta \ge 0
\end{aligned} \tag{25}$$

Unlike the parametric model (23), the above minmax formulation (25) gives continuous changes on the location of the Pareto-optimal point for continuous changes of the parameter t.

Thus, the optimal solution of (25) responds accurately to any given set of weights that gives priority to one stage over the other. In this sense, the unweighted minmax model (24) alignes more effectively with the notion of "neutrality" in the efficiency assessments than model (21) does and provides, thus, more balanced efficiency scores for the two stages.

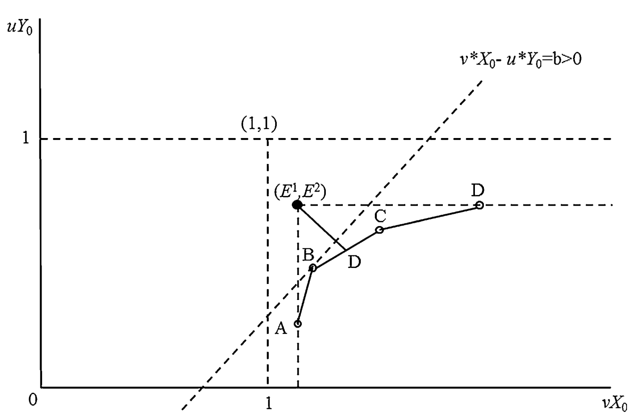

Fig. 3 The Pareto front of MOLP (20) and the optimal solution of model (24)

3.1.1 Aggregation of the individual efficiencies

As noticed in Liang et al. (2008), it is reasonable to define the overall efficiency of the two-stage process as the average (arithmetic mean) of the efficiencies of the two individual stages.

Also, Cook and Hababou (2001), although they did not directly address the issue of an "aggregate" measure of efficiency, they argued that this aggregate measure (overall efficiency) should be some average of the component scores. In this line of thought, the overall efficiency of unit j_0 is defined as:

$$\hat{e}^o_{j_0} = \frac{1}{2}(\hat{e}^1_{j_0} + \hat{e}^2_{j_0})$$

As the stage efficiencies are assumption-free, i.e. their assessment does not depend on any a priori definition of the overall efficiency, alternatively, they can be aggregated multiplicatively to get the overall efficiency as follows:

$$\hat{e}^o_{j_0} = \hat{e}^1_{j_0} \cdot \hat{e}^2_{j_0} = \frac{1}{v^* X_{j_0}} \cdot u^* Y_{j_0} = \frac{u^* Y_{j_0}}{v^* X_{j_0}}$$

In the next section, we compare the stage efficiencies and the overall efficiencies obtained by our approach with those obtained by the additive and the multiplicative methods presented in the previous section. Although the overall efficiency scores $\hat{e}^0$ and e^0 obtained respectively by our aggregation method (simple arithmetic average) and the additive decomposition model (13) are not comparable, because of the endogenous weights assumed for the two stages in the latter, in the case of the multiplicative decomposition model (7) the following hold:

Theorem 1 *If $\hat{e}^o_{j_0} = \hat{e}^1_{j_0} \cdot \hat{e}^2_{j_0}$ is the overall efficiency score of the evaluated unit j_0, with $\hat{e}^1_{j_0}, \hat{e}^2_{j_0}$ as derived by model (21), and $e^o_{j_0}$ is its overall efficiency score obtained from model (7) then $\hat{e}^o_{j_0} \le e^0_{j_0}$.*

Proof Let (v', w', u') be an optimal solution of model (7) with $e^0_{j_0} = u' Y_{j_0}$ and (v^*, w^*, u^*) an optimal solution of model (21) with $\hat{e}^o_{j_0} = u^* Y_{j_0} / v^* X_{j_0}$.

The following hold:

(a) (v', w', u') is an optimal solution in model (6). This is a direct implication of the C–C transformation.

(b) (v^*, w^*, u^*) is a feasible solution in (6). Indeed, (v^*, w^*, u^*) is optimal in the following ratio model:

$$\begin{aligned}
&\min\ \frac{\eta X_{j_0} - \omega Y_{j_0}}{\varphi Z_{j_0}}\\
&s.t.\\
&\varphi Z_j - \eta X_j \le 0, \quad j = 1,\ldots,n\\
&\omega Y_j - \varphi Z_j \le 0, \quad j = 1,\ldots,n\\
&\eta \ge 0, \varphi \ge 0, \omega \ge 0
\end{aligned}$$

which derives from (21) by applying the inverse C–C transformation: $\eta = v/t$, $\varphi = w/t$, $\omega = u/t$ with $t > 0$ such that $t\varphi\, Z_0 = 1$. As the above model and model (6) have the same feasible regions, (v^*, w^*, u^*) is feasible in (6). From (a) and (b) derives that $\hat{e}^o_{j_0} = u^* Y_{j_0} / v^* X_{j_0} \le u' Y_{j_0} = e^0_{j_0}$, which completes the proof.

Theorem 2 *If $\hat{e}^o_{j_0} = \hat{e}^1_{j_0} \cdot \hat{e}^2_{j_0}$ is the overall efficiency score of the evaluated unit j_0, with $\hat{e}^1_{j_0}, \hat{e}^2_{j_0}$ as derived by model (24), and $e^o_{j_0}$ is its overall efficiency score obtained from model (7) then $\hat{e}^o_{j_0} \le e^0_{j_0}$.*

Proof Let (v', w', u') be an optimal solution of model (7) with $e^0_{j_0} = u' Y_{j_0}$ and $(v^*, w^*, u^*, \delta^*)$ an optimal solution of model (24) with $\hat{e}^o_{j_0} = u^* Y_{j_0} / v^* X_{j_0}$. The following hold:

(a) The sub-vector (v^*, w^*, u^*) is a feasible solution of model (21). Indeed, given the optimal δ^*, the optimal sub-vector (v^*, w^*, u^*) satisfies the three last constraints of (24), which define the feasible region of (21).

(b) (v^*, w^*, u^*) is a feasible solution in (6) as well. The proof is as in Theorem 1 b).

Given (a) and (b), $\hat{e}^o_{j_0} \le e^0_{j_0}$ is direct implication of theorem 1.

3.2 Variable returns to scale

Our approach enables us to extend our developments under the variable returns-to-scale (VRS) assumption. Indeed, the VRS variant of model (21) is straightforward: when considering the VRS variants of models (15) and (16), as follows:

$$\begin{aligned}
&\min\ vX_{j_0} - d_1 - uY_{j_0} + d_2\\
&s.t.\\
&wZ_{j_0} = 1\\
&wZ_j - vX_j + d_1 \le 0, \quad j = 1, \ldots, n\\
&uY_j - wZ_j - d_2 \le 0, \quad j = 1, \ldots, n\\
&v \ge 0, w \ge 0, u \ge 0\\
&d_1, d_2 \quad \textit{free}
\end{aligned} \tag{26}$$

The additive decomposition approach of Chen et al. (2009) is extendable to VRS situations as well. However, this does not hold for the multiplicative model of Kao and Hwang (2008). Later, Kao and Hwang (2011) proposed an approach to decompose technical and scale efficiencies. Notably however, the principle that the VRS efficiency scores are not less than their CRS counterparts does not generally hold in neither the additive model or in our model (26) above. This irregularity can be attributed to the conflicting nature of the intermediate measures, which have different interpretations in the two stages. Adding, however, the constraints $vX_{j_0} - d_1 \le 1/\hat{e}^1_{CRS}$ and $uY_{j_0} - d_2 \ge \hat{e}^2_{CRS}$ in model (26), where $\hat{e}^1_{CRS}$ and $\hat{e}^2_{CRS}$ are the CRS efficiency scores obtained by model (21), rectifies this irregularity for the units where it is observed, without affecting the efficiency scores of the other units.

4 Illustration and experimentation

We apply our approach to the 24 Taiwanese non-life insurance companies originally studied in Kao and Hwang (2008). The authors consider a two-stage production process with two inputs (Operation expenses-X1 and Insurance expenses-X2), two intermediate measures (Direct written premiums-Z1 and Reinsurance premiums-Z2) and two final outputs (Underwriting profit-Y1 and Investment profit-Y2). Table 1 exhibits the data set.

Table 2 (columns 2–5) displays the efficiency scores obtained by applying our model (21) on the data of Table 1, and those obtained by model (24) (columns 8–12). Columns 6–7 present the ideal values of vX_{j_0} and uY_{j_0} in the bi-objective LP (20).

For comparison purposes, we give in Table 3 the results obtained from the additive decomposition model (13) of Chen et al. (2009) along with the weights (columns 2–6) and the corresponding results obtained from the multiplicative decomposition model (7) of Kao and Hwang (2008) (columns 7–9).

Although one can spot only a few differences among the individual efficiency scores obtained by model (21) and those obtained by models (13) and (7), in general, our approach does not yield the same efficiency scores for the individual stages with the other two methods. For instance, the stage-1 and stage-2 efficiency scores for DMU 16 (Allianz President) differ substantially from those obtained from the additive decomposition method. As regards the results obtained from the multiplicative decomposition method, the individual efficiency scores are different for 9 of the 24 units. Our experiments with different randomly generated data sets (100 data sets drawn from a uniform distribution, with 50 DMUs, 2 external inputs, 3 intermediate measures and 2 final outputs) revealed significant differentiation in the efficiency results between the three methods.

Figure 4 depicts the percentage of units in each run that showed different stage efficiency scores, with respect to model (21) and the additive model (13). The range of differences varies from 0 to 82 %. In only one case the efficiency scores were identical for all the units.

Table 1 Taiwanese non-life insurance companies data set (*source*: Kao and Hwang 2008)

DMU		X1	X2	Z1	Z2	Y1	Y2
1	Taiwan fire	1,178,744	673,512	7,451,757	856,735	984,143	681,687
2	Chung Kuo	1,381,822	1,352,755	10,020,274	1,812,894	1,228,502	834,754
3	Tai Ping	1,177,494	592,790	4,776,548	560,244	293,613	658,428
4	China mariners	601,320	594,259	3,174,851	371,863	248,709	177,331
5	Fubon	6,699,063	3,531,614	37,392,862	1,753,794	7,851,229	3,925,272
6	Zurich	2,627,707	668,363	9,747,908	952,326	1,713,598	415,058
7	Taian	1,942,833	1,443,100	10,685,457	643,412	2,239,593	439,039
8	Ming Tai	3,789,001	1,873,530	17,267,266	1,134,600	3,899,530	622,868
9	Central	1,567,746	950,432	11,473,162	546,337	1,043,778	264,098
10	The First	1,303,249	1,298,470	8,210,389	504,528	1,697,941	554,806
11	Kuo Hua	1,962,448	672,414	7,222,378	643,178	1,486,014	18,259
12	Union	2,592,790	650,952	9,434,406	1,118,489	1,574,191	909,295
13	Shingkong	2,609,941	1,368,802	13,921,464	811,343	3,609,236	223,047
14	South China	1,396,002	988,888	7,396,396	465,509	1,401,200	332,283
15	Cathay century	2,184,944	651,063	10,422,297	749,893	3,355,197	555,482
16	Allianz president	1,211,716	415,071	5,606,013	402,881	854,054	197,947
17	Newa	1,453,797	1,085,019	7,695,461	342,489	3,144,484	371,984
18	AIU	757,515	547,997	3,631,484	995,620	692,731	163,927
19	North America	159,422	182,338	1,141,950	483,291	519,121	46,857
20	Federal	145,442	53,518	316,829	131,920	355,624	26,537
21	Royal & Sunalliance	84,171	26,224	225,888	40,542	51,950	6,491
22	Aisa	15,993	10,502	52,063	14,574	82,141	4,181
23	AXA	54,693	28,408	245,910	49,864	0.1	18,980
24	Mitsui sumitomo	163,297	235,094	476,419	644,816	142,370	16,976

Analogously, Fig. 5 depicts the percentage of units in each run that showed different individual efficiency scores, with respect to model (21) and the multiplicative model (7). The range of differences varies from 23 to 97 %. None case was spotted with identical efficiency scores for all the units.

For the scores obtained from model (21), one can see that $\hat{e}^1 \geq e^1$ and $\hat{e}^2 \leq e^2$ where e^1 and e^2 are the stage-1 and stage-2 efficiency scores derived by either the additive or the multiplicative models. These relations are completely verified throughout our experiments mentioned above. As concerns the additive decomposition model (13), it is empirical evidence that the efficiency assessments are biased in favor of the second stage. As noted earlier, in reference to the results obtained by models (21) and (13), all the units but one (DMU 16) show identical individual scores for the two stages. A rigorous justification of both the similarities and the dissimilarities in the results can be given by solving model (23) for different values of the parameter $t;\ 0 < t < 1$. Table 4 exhibits, for a limited number of DMUs, the different efficiency scores with the indifference ranges of the parameter t. Due to space limitations, we have omitted most of the DMUs that show identical results for all the models. Column two shows the indifference ranges of the parameter t, within which the efficiency scores remain the same. Columns four and five present the efficiency scores for the two stages supported by the corresponding t-range in line. These scores correspond to successive extreme points (vertices) on the Pareto front generated by model (23). The asterisks in the last three columns indicate, among the alternative efficiency scores, those derived by the additive decomposition model (13) of Chen et al. (2009), our model (21) and the multiplicative model (7) of Kao and Hwang (2008), respectively. Column three depicts the endogenous weight t^2 assumed for the second stage in model (13). For the reasons explained at the end of Sect. 2 and since (24) is a composition rather than a decomposition model, the effect of changing the parameter t is strictly interpreted in relation to the weight t^2. The coinciding efficiency scores derived by models (13) and (21), for all the units but one (DMU 16) can now be rigorously justified by the fact that the supporting t-ranges contain both the weight values for t^2 assumed by model (13) as well as $t = 0.5$, which reflects the neutral (unweighted) character of model (21). As concerns the DMU 16, the t-range supporting the efficiency

Table 2 Results from our models (21) and (24)

DMU	Model (21)				Model (24)						
	$\hat{e}^1$	$\hat{e}^2$	$\hat{e}^o = (\hat{e}^1 + \hat{e}^2)/2$	$\hat{e}^o = \hat{e}^1 \cdot \hat{e}^2$	E_1	E_2	δ	$\hat{e}^1$	$\hat{e}^2$	$\hat{e}^o = (\hat{e}^1 + \hat{e}^2)/2$	$\hat{e}^o = \hat{e}^1 \cdot \hat{e}^2$
1	0.9926	0.7045	0.8485	0.6992	1.0075	0.7134	0.0079	0.9848	0.7054	0.8451	0.6947
2	0.9985	0.6257	0.8121	0.6248	1.0015	0.6275	0.0014	0.9971	0.6260	0.8116	0.6242
3	0.6900	1	0.8450	0.6900	1.4492	1	0	0.6900	1	0.8450	0.6900
4	0.7243	0.4200	0.5722	0.3042	1.3805	0.4323	0.0121	0.7181	0.4202	0.5692	0.3018
5	0.8307	0.9233	0.8770	0.7670	1.1940	1	0.0543	0.8011	0.9457	0.8734	0.7577
6	0.9606	0.4057	0.6831	0.3897	1.0377	0.4057	0.0019	0.9619	0.4037	0.6828	0.3883
7	0.7521	0.3522	0.5521	0.2649	1.3296	0.5378	0.1352	0.6827	0.4026	0.5426	0.2748
8	0.7256	0.3780	0.5518	0.2743	1.3782	0.5113	0.1038	0.6748	0.4076	0.5412	0.2750
9	1	0.2233	0.6116	0.2233	1	0.2920	0.0597	0.9437	0.2323	0.5880	0.2192
10	0.8615	0.5408	0.7012	0.4660	1.1607	0.6736	0.1139	0.7845	0.5597	0.6721	0.4391
11	0.7292	0.2066	0.4679	0.1507	1.3504	0.3267	0.0991	0.6899	0.2276	0.4587	0.1570
12	1	0.7596	0.8798	0.7596	1	0.7596	0	1	0.7596	0.8798	0.7596
13	0.8107	0.2431	0.5269	0.1970	1.2335	0.5435	0.2383	0.6794	0.3052	0.4923	0.2073
14	0.7246	0.3740	0.5493	0.2710	1.3800	0.5178	0.0956	0.6777	0.4222	0.5500	0.2861
15	1	0.6138	0.8069	0.6138	1	0.7047	0.0671	0.9371	0.6376	0.7874	0.5976
16	0.9072	0.3356	0.6214	0.3044	1.1023	0.3847	0.0250	0.8871	0.3597	0.6234	0.3191
17	0.7232	0.4597	0.5914	0.3325	1.3825	1	0.3817	0.5668	0.6183	0.5925	0.3504
18	0.7935	0.3262	0.5599	0.2588	1.2602	0.3737	0.0401	0.7691	0.3335	0.5513	0.2565
19	1	0.4112	0.7056	0.4112	1	0.4158	0.0038	0.9962	0.4120	0.7041	0.4104
20	0.9332	0.5857	0.7594	0.5465	1.0716	0.9014	0.2251	0.7712	0.6763	0.7238	0.5216
21	0.7505	0.2623	0.5064	0.1969	1.3324	0.2795	0.0127	0.7434	0.2668	0.5051	0.1984
22	0.5895	1	0.7948	0.5895	1.6963	1	0	0.5895	1	0.7948	0.5895
23	0.8426	0.4989	0.6707	0.4203	1.1764	0.5599	0.0520	0.8141	0.5079	0.6610	0.4135
24	1	0.0870	0.5435	0.0870	1	0.3351	0.2096	0.8267	0.1255	0.4761	0.1037

scores obtained by model (13) does not include the parameter value $t = 0.5$. This is exactly the source of differentiation in the results for DMU 16. Further, Table 3 shows that the parametric version of our model (21) can effectively locate the individual efficiency scores obtained from both the additive and the multiplicative decomposition methods.

Concerning the model (13), the fact that the weight given to the second stage is always at least as much as the weight given to the first stage, i.e. $t_j^2 \leq t_j^1$, the additive decomposition method biases the efficiency assessments for the individual stages. Thus, the overall efficiency score is biased as well. Indeed, although each DMU is free to choose its own multipliers so as to maximize its efficiency score, the freedom in selecting the weights t^1 and t^2 is structurally limited by $t_j^2 \leq t_j^1$. The case of unit #3 in Tables 2 and 3 is indicative. By selecting the weights $t^1 = 0.592$ and $t^2 = 0.408$ for the two stages, the stage efficiencies and the overall efficiency score obtained by the additive decomposition method are respectively $e^1 = 0.690$, $e^2 = 1$ and $e^o = 0.817$ (=0.592 × 0.690 + 0.408 × 1). We get the same stage efficiencies by our model (21). This is due to the fact that these scores are maintained for any value of the parameter $t \in (0,1)$ in model (23) (see Table 4). However, taking the simple (unweighted) average of the same individual scores gives an overall efficiency score 0.845, which is greater than the optimal overall efficiency obtained by the additive decomposition method. The same holds for units #5 and #22.

As concerns the results obtained by the minmax model (24), one can see that the efficiency scores of the individual stages are more balanced than those obtained by all the other models. The fact that three units, namely units 3, 12 and 22, show identical efficiency scores for the two stages across all models is justified by the fact that, for these units, the ideal point is attainable and thus the Pareto front degenerates in this single point. Figure 6 depicts the Pareto front ABDE generated by model (23) for unit 11, the Pareto-optimal point B (1.3713,0.2066) derived from model (21) that gives the optimal stage efficiency scores (0.7292,0.2066) as well as the Pareto-optimal point C (1.4495,0.2276) derived by the model (24) that gives the unique optimal stage efficiencies (0.6899, 0.2276).

Table 3 Results from models (13) and (7)

DMU	Chen et al. (2009)—model (13)					Kao and Hwang (2008)—model (7)		
	e^1	e^2	E^o	t^1	t^2	e^1	e^2	e^o
1	0.9926	0.7045	0.8491	0.502	0.498	0.9926	0.7045	0.6992
2	0.9985	0.6257	0.8122	0.500	0.500	0.9985	0.6257	0.6248
3	0.6900	1	0.8166	0.592	0.408	0.6900	1	0.6900
4	0.7243	0.4200	0.5965	0.580	0.420	0.7243	0.4200	0.3042
5	0.8307	0.9233	0.8727	0.546	0.454	0.8307	0.9233	0.7670
6	0.9606	0.4057	0.6887	0.510	0.490	0.9606	0.4057	0.3897
7	0.7521	0.3522	0.5804	0.571	0.429	0.6706	0.4124	0.2766
8	0.7256	0.3780	0.5795	0.580	0.420	0.6630	0.4150	0.2752
9	1	0.2233	0.6116	0.500	0.500	1	0.2233	0.2233
10	0.8615	0.5408	0.7131	0.537	0.463	0.8615	0.5408	0.4660
11	0.7291	0.2068	0.5088	0.578	0.422	0.6468	0.2534	0.1639
12	1	0.7596	0.8798	0.500	0.500	1	0.7596	0.7596
13	0.8107	0.2431	0.5565	0.552	0.448	0.6720	0.3093	0.2078
14	0.7246	0.3740	0.5773	0.580	0.420	0.6699	0.4309	0.2886
15	1	0.6138	0.8069	0.500	0.500	1	0.6138	0.6138
16	0.8856	0.3615	0.6395	0.530	0.470	0.8856	0.3615	0.3202
17	0.7232	0.4597	0.6126	0.580	0.420	0.6276	0.5736	0.3600
18	0.7935	0.3262	0.5868	0.558	0.442	0.7935	0.3262	0.2588
19	1	0.4112	0.7056	0.500	0.500	1	0.4112	0.4112
20	0.9332	0.5857	0.7654	0.517	0.483	0.9332	0.5857	0.5465
21	0.7505	0.2623	0.5412	0.571	0.429	0.7321	0.2743	0.2008
22	0.5895	1	0.7418	0.629	0.371	0.5895	1	0.5895
23	0.8426	0.4989	0.6854	0.543	0.457	0.8426	0.4989	0.4203
24	1	0.0870	0.5435	0.500	0.500	0.4287	0.3145	0.1348

5 Deriving the efficient frontier

A peculiarity of the two-stage DEA models, resulting from the conflicting nature of the intermediate measures, is that they are not capable of providing sufficient information to derive the efficient frontier, as it is with the standard DEA models. Chen et al. (2010) observed that the usual procedure of adjusting the inputs and outputs by the efficiency scores is not sufficient to yield a frontier projection neither in the additive nor in the multiplicative decomposition models. To overcome this inability, they proposed an envelopment model to locate the efficient frontier in the Kao and Hwang's (2008) multiplicative framework, by setting the intermediate measures as variables to be estimated. This approach enabled them to compute new levels for the inputs, the outputs and the intermediate measures that constitute efficient projections. These projections depend on the orientation selected. Actually, if an input orientation is assumed, new levels of inputs and intermediate measures are computed, while the original levels of outputs are maintained. Accordingly, assuming an output orientation, new levels of outputs and intermediate measures are obtained that maintain the original input levels. However, the levels of intermediate measures in these two cases differ substantially. Unfortunately, this technique cannot be applied in the additive decomposition framework. Chen et al. (2013) pointed out that the envelopment and the multiplier forms are two types of network DEA models, which use different concepts of efficiency; the former is developed explicitly on the basis of the production possibility set whereas the latter under the standard DEA ratio efficiency. Unlike the standard DEA, network DEA duality may not lead to a particular pair of network multiplier and envelopment models. Hence, Chen et al. (2010, 2013) proposed that the multiplier models should be used only for estimating the efficiency scores, whilst modified envelopment forms should be used for determining the frontier projections of the inefficient DMUs.

In the following, we formulate the envelopment form of model (21) and we use it as the basis to derive the efficient frontier of the two-stage process. To this end, consider the following model:

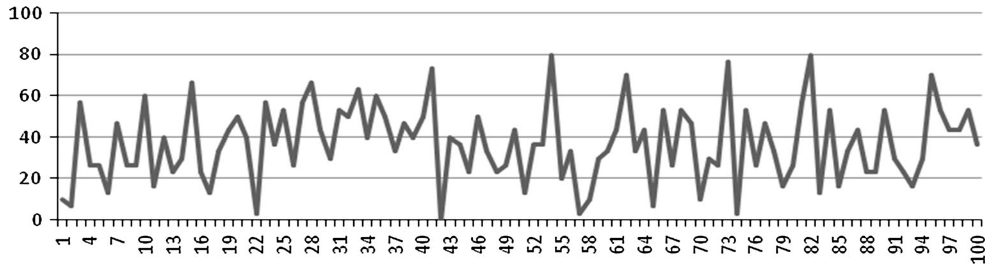

Fig. 4 Percentage of units showing different stage efficiencies: model (21) versus model (13)

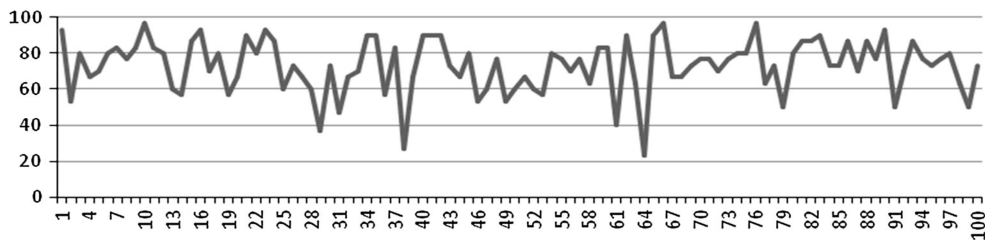

Fig. 5 Percentage of units showing different stage efficiencies: model (21) versus model (7)

$$\begin{aligned}
&g_{j_0} = \min vX_{j_0} - uY_{j_0} \\
&s.t. \\
&wZ_{j_0} = 1 \\
&-\hat{w}Z_{j_0} = -1 \\
&-wZ_j + vX_j \geq 0, \quad j = 1, \ldots, n \\
&-uY_j + \hat{w}Z_j \geq 0, \quad j = 1, \ldots, n \\
&w - \hat{w} = 0 \\
&v \geq 0, w \geq 0, \hat{w} \geq 0, u \geq 0
\end{aligned} \tag{27}$$

Model (27) is strictly equivalent to model (21). The difference in the formulation is that, in (27), different weight variables are used for the intermediate measures in the first and the second stage, which then are explicitly equalized in the last constraint. The dual of (27) is as follows:

$$\begin{aligned}
&\max \theta_1 - \theta_2 \\
&s.t. \\
&X\lambda + s^- = X_0 \\
&Y\mu - s^+ = Y_0 \\
&Z\lambda \geq \theta_1 Z_0 + a \\
&Z\mu \leq \theta_2 Z_0 + a \\
&\lambda \geq 0, \mu \geq 0, s^+ \geq 0, s^- \geq 0
\end{aligned} \tag{28}$$

where θ_1 and θ_2 are free in sign scalar variables and $a = (a_1, \ldots, a_q)$ is a vector of free in sign variables. In the optimal solution of (28), it is $\theta_1^* - \theta_2^* = g_{j_0}^* \geq 0$, where $g_{j_0}^*$ denotes the optimal objective value in (27), or equivalently, in model (21). If $\theta_1^* - \theta_2^* = g_{j_0}^* = 0$, then the evaluated unit is overall efficient. If $\theta_1^* - \theta_2^* = g_{j_0}^* > 0$, the unit it is overall inefficient. The interpretation of model (28) is straightforward if we take into account the double and conflicting role of the intermediate measures and the way that the primal model (21) was derived. With respect to the overall efficiency, whose components vX_{j_0}, uY_{j_0} appear in the objective function of (21) and (27), the model is of the non-oriented additive form and is capable of discriminating among overall efficient and overall inefficient units. With respect to the individual stages, the model simultaneously encompasses an output orientation for stage-1, expressed by the constraint $Z\lambda \geq \theta_1 Z_0 + a$, and an input orientation for the stage-2, expressed by the constraint $Z\mu \leq \theta_2 Z_0 + a$. Model (28) provides a dichotomic characterization of overall efficiency of the evaluated unit but not the individual efficiency scores. This limitation, however, is in analogy with the relevant limitation of Chen's et al. (2010) oriented envelopment model developed for the multiplicative decomposition method. Indeed, both the Chen's et al. (2010) model and our model (28) they provide the overall efficiency characterization they are structurally designed for, i.e. the former provides the overall efficiency score, as it is based on an oriented formulation, whereas the latter provides the overall efficiency status (efficient or inefficient) of the units being evaluated, as it is based on an non-oriented additive formulation with respect to external inputs and the final outputs. The analogy is completed by the fact that none of the above provides the efficiency scores for the individual stages. Although $\theta_1^* = 1/\hat{e}^1, \theta_2^* = \hat{e}^2$ are feasible, yet optimal values of the variables θ_1 and θ_2, it is unlikely that they will be obtained by solving (28). In fact, in the optimal solution $(\lambda^*, \mu^*, \theta_1^*, \theta_2^*, a^*)$ of (28), $\theta_1^*$and θ_2^* can take any values,

Table 4 Efficiency scores obtained by model (23) for different values of the parameter *t*

DMU	*t* (indifference ranges)	t^2	e^1	e^2	Model (13)	Model (21)	Model (7)
3	(0,1)	0.408	0.690	1	*	*	*
5	(0, 0.3355)		0.738	1			
	[0.3355, 0.9228)	0.454	0.831	0.923	*	*	*
	[0.9228, 1)		0.837	0.806			
7	(0, 0.048)		0.300	0.538			
	[0.048, 0.0528)		0.382	0.502			
	[0.0528, 0.0575)		0.514	0.464			
	[0.0575, 0.1368)		0.575	0.452			
	[0.1368, 0.2718)		0.671	0.412			*
	[0.2718, 1)	0.429	0.752	0.352	*	*	
8	(0, 0.0702)		0.390	0.511			
	[0.0702, 0.0907)		0.491	0.472			
	[0.0907, 0.1192)		0.619	0.430			
	[0.1192, 0.2215)		0.663	0.415			*
	[0.2215, 1)	0.420	0.726	0.378	*	*	
11	(0, 0.1133)		0.472	0.327			
	[0.1133, 0.2114)		0.647	0.253			*
	[0.2114, 0.651)	0.422	0.729	0.207	*	*	
	[0.651, 1)		0.741	0.168			
13	(0, 0.1148)		0.338	0.543			
	[0.1148, 0.1355)		0.405	0.480			
	[0.1355, 0.1647)		0.519	0.395			
	[0.1647, 0.2007)		0.672	0.309			*
	[0.2007, 0.211)		0.729	0.280			
	[0.211, 1)	0.448	0.811	0.243	*	*	
14	(0, 0.0298)		0.310	0.518			
	[0.0298, 0.0334)		0.392	0.497			
	[0.0334, 0.0371)		0.521	0.475			
	[0.0371, 0.1367)		0.579	0.468			
	[0.1367, 0.3356)		0.670	0.431			*
	[0.3356, 1)		0.725	0.374	*	*	
16	(0, 0.0281)		0.599	0.385			
	[0.0281, 0.0504)		0.744	0.375			
	[0.0504, 0.1406)		0.869	0.365			
	[0.1406, 0.491)	0.470	0.886	0.362	*		*
	[0.491, 1)		0.907	0.336		*	
17	(0, 0.1358)		0.251	1			
	[0.1358, 0.1461)		0.333	0.845			
	[0.1461, 0.1564)		0.466	0.698			
	[0.1564, 0.2071)		0.529	0.651			
	[0.2071, 0.3511)		0.628	0.574			*
	[0.3511, 0.9451)	0.420	0.723	0.460	*	*	
	[0.9451, 1)		0.723	0.455			
21	(0, 0.0619)		0.692	0.280			
	[0.0619, 0.2625)		0.732	0.274			*
	[0.2625, 1)	0.429	0.751	0.262	*	*	
24	(0, 0.1051)		0.399	0.335			
	[0.1051, 0.1441)		0.429	0.314			*
	[0.1441, 0.1663)		0.908	0.107			
	[0.1663, 1)	0.500	1.000	0.087	*	*	

such that $\theta_1^* - \theta_2^* = g_{j_0}^*$, by adjusting accordingly the values of a^*. This is because the variables θ_1, θ_2 and a are free in sign and unbounded. Thus, the optimal λ^* and μ^* as well as the optimal value of the objective function are not affected if we require $\theta_1 \geq 1$ and $\theta_2 \leq 1$, which reflect the output and the input orientation assumed for stage 1 and stage 2 respectively.

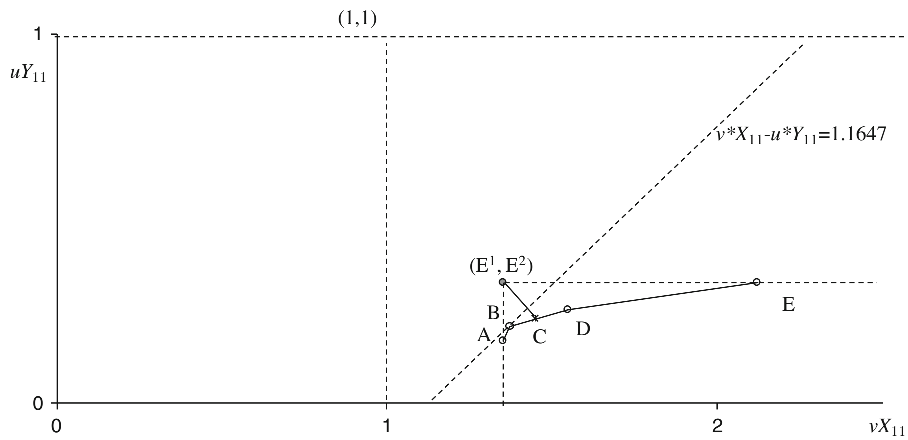

Fig. 6 The Pareto front of unit 11 and the Pareto points derived by models (21) and (24)

Another issue with model (28) is that the divergent orientations imposed by the constraints $Z\lambda \geq \theta_1 Z_0 + a$ and $Z\mu \leq \theta_2 Z_0 + a$ on the intermediate measures do not allow it to provide correct projections of the inefficient units on the efficient frontier. Chen et al. (2010) overcame an analogous issue in their developments by solving a modified model, where the observed values of the intermediate measures Z_0 in the constraints $Z\lambda \geq Z_0$ and $Z\mu \leq Z_0$ are replaced with variables$\tilde{Z}_0$ that represent the projections for the intermediate measures. Actually, this is the implementation of the "free link" case, as this sort of modification is characterized in Tone and Tsutsui (2009). The transition of our basic envelopment model (28), in a form capable to derive the efficient frontier, has exactly the same rationale: to make the right hand sides of the above two constraints coincide. Setting $\theta_1 = \theta_2 = 1$, i.e. at the value where $\theta_1 \geq 1$ and $\theta_2 \leq 1$ meet, the right hand sides of the last two constraints in (28) become $Z_0 + a = \tilde{Z}_0$, where the variables $\tilde{Z}_0$ represent, as in Chen et al. (2010), the targets for the intermediate measures. Hence, the following model is solved to obtain the projections of the inefficient units on the frontier:

$$\begin{aligned} &\max\ es^- + es^+ \\ &s.t. \\ &X\lambda + s^- = X_0 \\ &Y\mu - s^+ = Y_0 \\ &Z\lambda \geq \tilde{Z}_0 \\ &Z\mu \leq \tilde{Z}_0 \\ &\lambda \geq 0, \mu \geq 0, s^+ \geq 0, s^- \geq 0 \end{aligned} \tag{29}$$

where the variables $\tilde{Z}_0$ are left free in sign. Actually, as $\tilde{Z}_0$ will never take negative values because of the last constraint, the natural restrictions $\tilde{Z}_0 \geq 0$ are redundant and, thus, omitted. Once an optimal solution $(\lambda^*, \mu^*, \tilde{Z}_0^*, s^{-*}, s^{+*})$ of model (29) is obtained, the evaluated unit is overall efficient if $s^{+*} = s^{-*} = 0$. The efficient projections of the inefficient units are as follows:

$$\hat{X}_0 = X_0 - s^{-*}, \quad \hat{Y}_0 = Y_0 + s^{+*}, \quad \hat{Z}_0 = \tilde{Z}_0^*$$

Hence, an inefficient DMU (X_0, Z_0, Y_0) is projected onto the efficient frontier at the point $(\hat{X}_0, \hat{Z}_0, \hat{Y}_0)$. Model (29) is now in a pure additive form. Indeed, the dual of (29) is as follows:

$$\begin{aligned} &\min vX_{j_0} - uY_{j_0} \\ &s.t. \\ &-wZ_j + vX_j \geq 0, \quad j = 1, \ldots, n \\ &-uY_j + \hat{w}Z_j \geq 0, \quad j = 1, \ldots, n \\ &v \geq e \\ &u \geq e \\ &w - \hat{w} = 0 \\ &w \geq 0, \hat{w} \geq 0 \end{aligned} \tag{30}$$

or

$$\begin{aligned} &\min vX_{j_0} - uY_{j_0} \\ &s.t. \\ &-wZ_j + vX_j \geq 0, \quad j = 1, \ldots, n \\ &-uY_j + wZ_j \geq 0, \quad j = 1, \ldots, n \\ &v \geq e \\ &u \geq e \\ &w \geq 0 \end{aligned} \tag{31}$$

Table 5 shows the projections obtained by applying model (29).

The efficiency status of these projections is verified by applying models (21) and (24) to an expanded data set that contains both the original DMUs (Table 1) and their projections (Table 5). Indeed, the results showed that all the

Table 5 Projections for Taiwanese non-life insurance companies obtained by model (29)

DMU	X1	X2	Z1	Z2	Y1	Y2
1	1,178,744.00	673,512.00	6,574,261.72	1,405,802.31	6,321,322.37	681,687.00
2	1,381,822.00	1,282,484.97	9,984,108.87	2,701,675.41	14,883,490.16	834,754.00
3	1,177,494.00	592,790.00	6,301,206.75	1,173,093.25	5,222,222.92	658,428.00
4	601,320.00	566,085.63	4,342,422.69	1,215,574.75	6,851,140.78	348,724.99
5	6,699,063.00	3,531,614.00	36,299,271.58	7,179,671.93	30,095,254.05	3,925,272.00
6	1,160,508.24	668,363.00	5,711,178.98	1,598,730.82	9,010,659.25	458,645.09
7	1,942,833.00	1,443,100.00	11,869,185.17	3,322,542.01	18,726,288.14	953,173.33
8	3,253,093.02	1,873,530.00	16,009,361.92	4,481,502.04	25,258,340.80	1,285,656.65
9	1,567,746.00	950,432.00	8,068,966.39	2,258,746.45	12,730,595.02	647,990.87
10	1,303,249.00	1,226,885.07	9,411,391.65	2,634,531.66	14,848,570.42	755,796.41
11	1,167,542.17	672,414.00	5,745,794.88	1,608,420.85	9,065,273.56	461,424.97
12	2,563,321.18	650,952.00	9,356,387.29	1,127,326.43	6,005,636.18	909,295.00
13	2,376,711.46	1,368,802.00	11,696,448.21	3,274,187.74	18,453,757.03	939,301.42
14	1,396,002.00	988,888.00	8,245,849.34	2,308,261.30	13,009,667.34	662,195.73
15	1,467,228.98	651,063.00	6,454,920.94	1,456,357.91	8,126,418.64	555,482.00
16	720,706.14	415,071.00	3,546,792.34	992,853.88	5,595,856.36	284,830.66
17	1,453,797.00	1,085,019.00	8,910,486.74	2,494,313.31	14,058,281.15	715,570.46
18	757,515.00	547,997.00	4,545,666.61	1,272,468.84	7,171,803.41	365,046.81
19	159,422.00	150,080.66	1,151,263.40	322,273.26	1,816,374.92	92,453.99
20	92,925.67	53,518.00	457,312.68	128,015.58	721,512.80	36,725.20
21	45,533.89	26,224.00	224,084.75	62,728.06	353,543.70	17,995.47
22	15,993.00	10,502.00	88,313.63	24,721.64	139,334.46	7,092.16
23	49,326.07	28,408.00	242,747.09	67,952.21	382,987.70	19,494.18
24	163,297.00	153,728.61	1,179,246.65	330,106.62	1,860,524.74	94,701.23

projected units are rendered efficient in both stages as well as in the overall sense, while the efficiency scores of the original units remained unchanged. This confirms that our approach accurately determines the improvement targets on the efficient frontier. We extended our calculations by adding in the expanded data set the projections derived by the other approaches. Particularly, we incorporated the projections of Chen et al. (2010) and those obtained by the non-oriented envelopment network DEA model (Chen et al. 2013), which is a modification of the slacks based measure (SBM) model of Tone and Tsutsui (2009). The results showed that all the projections, no matter the method that they derive from, are efficient. This verifies that our models (21) and (24) maintain the efficiency status of alternative projections obtained by the other methods. Notably, our projections are deemed efficient as well, when tested with the models (7) and (13).

6 Conclusion

In this paper, we firstly introduced a novel approach to assess the individual and the overall efficiencies in two-stage DEA, under the common assumption of series relationship between the two stages. Our approach effectively overcomes the shortcomings highlighted for the additive and the multiplicative decomposition methods, by providing unique and unbiased efficiency scores for the two stages. Based on a reverse perspective in aggregating the individual efficiency scores, i.e. the composition as opposed to the decomposition approach, we estimate first the individual efficiencies for the two stages, which then can be aggregated in either an additive or a multiplicative form to obtain the overall efficiency. Our modeling approach is based on the selection of an output orientation for the first stage and an input orientation for the second stage, with respect to the standard DEA ratio models. In this manner, the intermediate measures are used as the basis to link the efficiency assessment models for the two stages in a single linear program. The proposed approach is straightforwardly extended to fit VRS situations. Acknowledging the inadequacies observed for the envelopment network DEA models, we presented a method to derive the efficient frontier in two-stage DEA, which stems from the envelopment form of our basic multiplier model.

Acknowledgments The authors wish to thank the two anonymous reviewers for their constructive comments, which helped us improve the quality of the paper. This research has been co-financed by the

European Union (European Social Fund–ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF)—Research Funding Program: THALES-Investing in knowledge society through the European Social Fund.

References

Banker RD, Charnes A, Cooper WW (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis. Manage Sci 30:1078–1092

Castelli L, Pesenti R, Ukovich W (2010) A classification of DEA models when the internal structure of the decision making units is considered. Ann Oper Res 173:207–235

Charnes A, Cooper WW (1962) Programming with linear fractional functional. Naval Res Logist 9:181–185

Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision making units. Eur J Oper Res 2:429–444

Chen Y, Zhu J (2004) Measuring information technology's indirect impact on firm performance. Inf Technol Manage 5:9–22

Chen Y, Cook WD, Li N, Zhu J (2009) Additive efficiency decomposition in two-stage DEA. Eur J Oper Res 196:1170–1176

Chen Y, Cook WD, Zhu J (2010) Deriving the DEA frontier for two-stage DEA processes. Eur J Oper Res 202:138–142

Chen Y, Cook WD, Kao C, Zhu J (2013) Network DEA pitfalls: divisional efficiency and frontier projection under general network structures. Eur J Oper Res 226:507–515

Cook WD, Hababou M (2001) Sales performance measurement in bank branches. Omega 29:299–307

Färe R, Grosskopf S (1996) Productivity and intermediate products: a frontier approach. Econ Lett 50:65–70

Geoffrion AM (1968) Proper efficiency and theory of vector maximization. J Math Anal Appl 22:618–630

Kao C, Hwang S-N (2008) Efficiency decomposition in two-stage data envelopment analysis: an application to non-life insurance companies in Taiwan. Eur J Oper Res 185:418–429

Kao C, Hwang S-N (2011) Decomposition of technical and scale efficiencies in two-stage production systems. Eur J Oper Res 211:515–519

Liang L, Cook WD, Zhu J (2008) DEA models for two-stage processes: game approach and efficiency decomposition. Naval Res Logist 55:643–653

Seiford LM, Zhu J (1999) Profitability and marketability of the top 55 US commercial banks. Manage Sci 45(9):1270–1288

Tone K, Tsutsui M (2009) Network DEA: a slacks-based measure approach. Eur J OperRes 197:243–252