Expert Systems With Applications 99 (2018) 213–230

Contents lists available at ScienceDirect

Expert Systems With Applications

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

A novel two-stage DEA production model with freely distributed initial inputs and shared intermediate outputs

Mohammad Izadikhah[a], Madjid Tavana[b,c,*], Debora Di Caprio[d,e], Francisco J. Santos-Arteaga[f]

[a] Department of Mathematics, College of Science, Arak Branch, Islamic Azad University, Arak, Iran
[b] Business Systems and Analytics Department, Distinguished Chair of Business Analytics, La Salle University, Philadelphia, PA 19141, USA
[c] Business Information Systems Department, Faculty of Business Administration and Economics, University of Paderborn, D-33098 Paderborn, Germany
[d] Department of Mathematics and Statistics, York University, Toronto, M3J 1P3, Canada
[e] Polo Tecnologico IISS G. Galilei, Via Cadorna 14, 39100, Bolzano, Italy
[f] Faculty of Economics and Management, Free University of Bolzano, Bolzano, Italy

ARTICLE INFO

Article history:
Received 21 June 2017
Revised 9 October 2017
Accepted 1 November 2017
Available online 10 November 2017

Keywords:
Data envelopment analysis
Intermediate measure
Efficiency
Shared flow
Two-stage process

ABSTRACT

Conventional data envelopment analysis (DEA) models treat the decision-making units (DMUs) as black-boxes: inputs enter the system and outputs exit the system, with no consideration for the intermediate steps characterizing the DMUs. As a result, intermediate measures are lost in the process of changing the inputs to outputs and it becomes difficult, if not impossible, to provide individual DMU managers with specific information on what part of a DMU is responsible for the overall inefficiency. This study defines a two-stage DEA model, where each DMU is composed of two sub-DMUs in series, the intermediate products by the sub-DMU in Stage 1 are partly consumed by the sub-DMU in Stage 2, and the initial inputs of the DMU can be freely allocated in both stages. Also, there are additional inputs directly consumed in Stage 2 while part of the outputs of Stage 1 are final outputs. We develop four new linear models to determine the upper and lower bounds of the efficiencies of the two sub-DMUs in a non-cooperative setting and a linear model to calculate the overall efficiency of DMU in a cooperative setting. That is, the overall efficiency of a DMU is modelled in a cooperative setting via upper and lower bounds obtained in the non-cooperative one. The proposed two-stage DEA method allows for important applications to several management areas. A case study in the banking industry is presented to demonstrate the applicability and exhibit the efficacy of the proposed models.

1. Introduction

Data envelopment analysis (DEA) is an effective non-parametric evaluation method for measuring the relative efficiency of a set of decision making units (DMUs) each of which uses multiple inputs to produce multiple outputs. In the traditional DEA methods, DMUs are treated as black-boxes, that is, the internal structure of the DMUs is often ignored. As a result, the focus of the investigation is on the single operational processes with a set of initial inputs and final outputs that is unable of pinpointing the sources of inefficiency within the DMUs (Lewis, Mallikarjun, & Sexton, 2013; Wang, Huang, Wu, & Liu, 2014). On the other hand, in the two-stage DEA models, DMUs are modelled as systems composed of two sub-DMUs in series where the outputs of the sub-DMU in Stage 1, known as intermediate measures/products/outputs, are considered as inputs of the sub-DMU in Stage 2. As a result, a two-stage DEA model allows one to further investigate the structure of a DMU and its processes and, hence, to identify the misallocation of the inputs among the sub-DMUs (Du, Liang, Chen, Cook, & Zhu, 2011; Ebrahimnejad, Tavana, Lotfi, Shahverdi, & Yousefpour, 2014; Yu, Shi, & Song, 2013).

Despite appearing as the simplest multi-stage approach to efficiency evaluation, two-stage DEA models are the building blocks for the study of series systems whose DMUs consist of multiple sub-DMUs operating through procedures of different complexity. The work of Charnes et al. (1986) on army recruitment using a two-stage approach was the first study to discuss the loss of information intrinsic to single-stage models. In recent years, many researchers have studied several applications of two-stage DEA models to diverse decision making situations such as healthcare man-

* Corresponding author at: Business Systems and Analytics Department, Distinguished Chair of Business Analytics, La Salle University, Philadelphia, PA 19141, USA.
E-mail addresses: m-izadikhah@iau-arak.ac.ir (M. Izadikhah), tavana@lasalle.edu (M. Tavana), dicaper@mathstat.yorku.ca, debora.dicaprio@istruzione.it (D. Di Caprio), fsantosarteaga@unibz.it (F.J. Santos-Arteaga).

https://doi.org/10.1016/j.eswa.2017.11.005
0957-4174/

agement programs (Schinnar, Kamis-Gould, Delucia, & Rothbard, 1990), education programs (Lovell, Walters, & Wood, 1994), and sport teams (Sexton & Lewis, 2003), among others. Among the many possible applications, those to the banking industry have been attracting a considerable amount of attention (Seiford & Zhu, 1999; Ebrahimnejad et al., 2014). More in general, Castelli, Pesenti, and Ukovich (2010) provide a comprehensive categorized overview of the models and methods that have been developed for different multi-stage production architectures.

The problem of evaluating the efficiency of DMUs can also be considered from the game theoretical perspective. In this sense, there are two main approaches to the problem: one approach models two-stage processes as non-cooperative games, the other as centralized systems aiming at maximizing their overall efficiency (Li, 2017). One of the first studies to employ the concepts of cooperative and non-cooperative games in an efficiency evaluation context referring to two-stage DMUs is due to Liang, Yang, Cook, and Zhu (2006). In the non-cooperative setting, one of the two stages is considered more important than the other: the former is identified with *the leader* while the latter with *the follower*. As for the efficiency scores, one starts by measuring and optimizing the leader's efficiency without considering the follower; hence, the follower's efficiency is measured under the constraint that the leader's efficiency remains constant. In the cooperative setting, both stages have the same importance and their efficiency scores must be measured and optimized at the same time (Hosseinzadeh Lotfi, Jahanshahloo, Hemati, & Givehchi, 2012; Mahdiloo et al., 2016). For example, interpreting the offices of a regional R&D program as two-stage DMUs whose first stage consists of Premium Acquisition and the second stage is Profit Generation, we could assume that the Stage 1 is the leader. Thus, the efficiency of Stage 2 would be computed subject to the constraint that the efficiency of Stage 1 remains fixed. Alternatively, we could assume that Stage 2 is the leader and, hence, measure the efficiency of Stage 1 keeping the efficiency of Stage 2 constant (Li, Chen, Liang, & Xie, 2012; Yu & Shi, 2014).

Kao and Hwang (2008) presented a two-stage DEA model where the efficiency of the entire process can be decomposed into the product of the efficiencies of its two sub-processes. However, their approach has a few limitations (Guo, Abbasi Shureshjani, Foroughi, & Zhu, 2017). To start, it can be applied only when the assumption of constant returns to scale is satisfied. Moreover, it allows to model only *closed* multi-stage processes, that is, processes where the outputs of a stage are entirely used as inputs in the following stage with no extra input added. If, for instance, additional inputs are introduced from one stage to the following one, the resulting model becomes non-linear in the multiplicative approach. This is one of the main drawbacks associated with two-stage structures. Zha and Liang (2010) slightly improved the situation by presenting a method for studying a two-stage production process in series where the initial inputs can be freely allocated in the two stages. However, their method yields a parametric model that has other pitfalls. In particular, it does not allow for additional direct inputs to be used in Stage 2 or the existence of shared intermediate products. In order to improve both the method of Zha and Liang (2010) and the structure of two-stage systems, Yu and Shi (2014) proposed a two-stage DEA model where additional inputs are allocated in Stage 2 and part of the outputs of Stage 1 are used as inputs in Stage 2. However, they formulated a parametric model to solve their two-stage structure that does not allow for shared inputs or final outputs to be produced directly in Stage 1. The method proposed in this study deals with the aforementioned problems providing a viable solution.

In this paper, we consider a novel two-stage DEA model, where each DMU is composed of two sub-DMUs in series, the intermediate products of the sub-DMU in Stage 1 are partly consumed by the sub-DMU in Stage 2, and the initial inputs of the DMU can be freely allocated to either one of the two sub-DMUs. Also, there are additional inputs directly consumed in Stage 2 while the sub-DMU in Stage 1 is allowed to produce final outputs.

Building on the concept of Stackelberg leader–follower game and the work of Yu and Shi (2014), we formulate suitable fractional programming problems that allow to determine the upper and lower bounds of the efficiencies of the two sub-DMUs in the non-cooperative setting. Afterwards, we consider the cooperative setting and define a fractional programming model to calculate the overall efficiency of DMUs. All the models are linearized by applying a Charnes–Cooper transformation (Charnes & Cooper, 1962).

To show the advantages of using the proposed method in place of the existing ones, we provide some numerical comparisons with the methods defined by Kao and Hwang (2008), Zha and Liang (2010) and Yu and Shi (2014). These comparisons show that our method is much more general than the existing models in terms of closeness and consistency of the solutions.

According to Zha and Liang (2010), Matthews (2013), and Akther, Fukuyama, and Weber (2013), a banking system can be viewed as a two-stage process. Moreover, following Seiford and Zhu (1999), Zhu (2000), Luo (2003), and Liu (2011b), the first stage can be used to measure the *profitability* of a financial company (its ability to generate revenue and profit using its labor, assets, and capital resources) while the second stage measures its *marketability* (the ability of the company to generate revenue and profit through its stock market performance). We will modify the two-stage evaluation framework proposed by Seiford and Zhu (1999) and Liu (2011b) and adapt it to analyze 15 branches of a commercial bank in the Philadelphia metropolitan area. The results obtained by using our intermediate-like approach are compared with those delivered by a classical black-box approach showing that the latter approach does not suffice to adequately evaluate the aggregate performances of the DMUs composing a system.

The paper unfolds as follows. Section 2 presents a literature review while Section 3 provides a detailed review of three specific two-stage DEA models that will be used as reference points to define our approach. Section 4 extends Yu and Shi (2014)'s method to the proposed two-stage DEA production model with freely distributed inputs and shared intermediate outputs. Section 5 provides some numerical comparisons among the proposed method and those defined by Kao and Hwang (2008), Zha and Liang (2010) and Yu and Shi (2014) showing the advantages of our model over the existing ones. Section 6 describes a case study in the US banking industry and discusses the corresponding numerical results. Finally, Section 7 concludes and suggests some future research directions.

2. Literature review

In this section, we review some the vast literature related to the present study.

2.1. A general review of two-stage DEA models

2.1.1. A glance at some early applications of two-stage DEA

Färe and Whittaker (1995) and Färe and Grosskopf (1996) applied an input oriented two-stage network DEA model to study relative efficiencies in dairy production processes. Wang, Gopal, and Zionts (1997) and Noulas, Hatzigayios, Lazaridis, and Lyroudi (2001) proposed applications of two-stage DEA to information technology (IT) and to non-life insurance policies, respectively. Färe and Grosskopf (2000) presented a network DEA model for the Swedish Institute for Health Economics. Seiford and Zhu (1999) divided a complicated production process into independent sub-processes and, hence, calculated the efficiencies of

the first stage, the second stage, and the whole process via three independent DEA models. Zhu (2000) applied the same two-stage modelling process to the companies of Fortune Global 500.

Luo (2003) measured the efficiencies of US commercial banks in terms of profitability and marketability via a two-stage DEA process showing that bank inefficiency is related to the marketability stage rather than to the profitability one. Chen and Zhu (2004) presented a two-stage process to measure the profitability and marketability of U.S. commercial banks. They evaluated the performance in terms of profitability considering labor and assets as inputs and profits and revenue as outputs. Hence, the efficiency in terms of marketability was calculated using the outputs obtained in the profitability stage as inputs to produce the final outputs determining the market value, that is, returns and earnings per share.

More recently, Chen, Liang, Yang, and Zhu (2006) developed a nonlinear DEA model to analyze both the impact of IT on a multiple-stage business operation and how to allocate the available IT resources to guarantee efficiency. They showed, in particular, that two-stage network DEA with a single intermediate output can be treated as a parametric linear model. Yang (2006) created a two-stage DEA model allowing for applications to the Canadian life and health insurance companies.

2.1.2. Decomposition approach for evaluating the overall efficiency

Kao and Hwang (2008) presented a two-stage DEA model where the efficiency of the whole process can be decomposed into the product of the efficiencies of its two sub-processes. Chen, Cook, Li, and Zhu (2009) generalized the two-stage DEA model presented by Kao and Hwang (2008) by implementing an additive efficiency decomposition method that allows to measure the overall efficiency scores as (weighted) sums of the single stage efficiencies. They also examined the relationships between the two-stage DEA approach of Chen and Zhu (2004) and the one of Kao and Hwang (2008). Chen, Cook, and Zhu (2010) developed a method capable of determining the frontier points of inefficient DMUs within the two-stage DEA framework. Wang and Chin (2010) focused on alternative DEA models for two-stage processes and managed both extending the model presented by Kao and Hwang (2008) so as to allow for the variable returns to scale (VRS) assumption and generalizing the model of Chen et al. (2009) so as to account for the relative importance weights of the single stages. Zha and Liang (2010) presented a method for studying a two-stage production process in series where the initial inputs can be freely allocated in the different stages. However, Kao and Hwang (2011) kept the focus on the efficiency assessment problem of closed two-stage systems, that is, systems where all the outputs from the first stage become the inputs of the second stage.

2.1.3. Game theoretical approach for evaluating the overall efficiency

Liang, Cook, and Zhu (2008) proposed a game theoretical approach to the efficiency assessments of two-stage processes. They also investigated the relationships among non-cooperative, centralized, and standard DEA approaches. Cook, Liang, and Zhu (2010) carried out a review of the main two-stage models existing in the literature discussing the relationships among the various approaches. They showed that all the existing methods are based on either the Stackelberg (leader-follower) game concept or the cooperative game approach. Among others, it is worth mentioning the work of Du et al. (2011), Hosseinzadeh Lotfi et al. (2012) and Du, Chen, Cook, Liang, and Zhu (2014) who developed a Nash bargaining game model to measure the performance of two-stage DMUs.

Li et al. (2012) extended the model of Liang et al. (2008) by allowing the inputs of the second stage to consist of outputs produced in the first stage as well as additional inputs directly supplied to the second stage. Yu and Shi (2014) proposed a two-stage DEA model where additional inputs are allocated in the second stage and part of the outputs of the first stage are considered as final outputs. They analyzed the non-cooperative setting to determine the upper and lower bounds of the efficiencies of the two sub-DMUs composing a DMU. Hence, they formulated a nonlinear programming problem to evaluate the overall cooperative efficiency and defined a parametric transformation to solve it.

2.1.4. Further developments and ongoing research lines

Liu (2011a) proposed a method that allows to compute the sub-process and overall efficiencies simultaneously while Liu (2011b) measured the profitability and marketability efficiencies of financial holding companies in Taiwan based on the series relationship existing between the individual stages.

Despotis, Koronakos, and Sotiros (2012) presented an alternative approach to additive efficiency decomposition in two-stage DEA. Their approach aggregates in a single linear program the first and second stage efficiency evaluation models using the intermediate measures as pivot. Akther et al. (2013) examined the performance of 21 Bangladesh banks by using a two-stage network Slacks-based inefficiency measuring procedure.

Lewis et al. (2013) presented an un-oriented two-stage DEA model to measure efficiency in situations where reducing input quantities and increasing output quantities are aimed for simultaneously. Yu et al. (2013) discussed the relationship between the models of Chen et al. (2009) and Wang and Chin (2010). Sun, Xiong, and Tang (2013) proposed a modified relational two-stage DEA model under the VRS assumption relaxing the requirement that all the outputs of the first stage must be inputs of the second stage.

Ebrahimnejad et al. (2014) proposed a three-stage DEA model with two independent parallel stages linking to a third final stage. Liu (2014b) defined a fuzzy two-stage DEA model accounting for fuzzy input–output data while restricting the weights to given ranges. Liu (2014a) worked out a methodology to evaluate fuzzy efficiencies in cases where it is not possible to determine the exact membership functions of overall efficiencies obtained through fuzzy two-stage procedures. Tavana and Khalili-Damghani (2014) proposed a two-stage fuzzy DEA model that measures the efficiency of DMUs and their sub-DMUs. Following a Stackelberg (leader–follower) game theoretical approach, they managed to sequentially decompose the efficiencies of each DMU via the efficiencies of its sub-DMUs.

Kao (2014) analyzed the efficiency assessment and decomposition problem for general multi-stage systems. The model assumed each stage to consume exogenous inputs and intermediate products (produced from the preceding stage) to produce final outputs and intermediate products (to use in the succeeding stage). Kao and Hwang (2014) developed a multi-period two-stage DEA methodology that considers the operations of individual periods. Their model allows measuring the overall and single period efficiencies at the same time, with the former efficiency expressed as a weighted average of the latter ones. At the same time, the decomposition of the overall efficiency of a two-stage system into the product of the efficiencies of the two stages allows to identify the sources of inefficiency in a DMU.

In order to locate the sources of efficiency and economies of scale of a network firm, Sahoo, Zhu, Tone, and Klemen (2014) suggested two approaches depending on whether or not the two-stage network technology structure considered in each approach allows for allocative inefficiency. The first approach assumed a single network technology for two interdependent sub-technologies. In the second approach, however, the technology structure was determined under the condition that its sub-technologies are independent, that is, assuming that there is no allocative inefficiency.

Liu, Zhou, Ma, Liu, and Shen (2015) defined two-stage DEA models with undesirable variables (inputs, intermediate measures, outputs) by employing the free disposal axioms. The two-stage DEA model proposed by Ma (2015) allows to consider inputs and intermediate measures simultaneously for both efficiency evaluations and decompositions. The relational linear DEA model presented by Toloo, Emrouznejad, and Moreno (2015) can be used to assess the efficiency of two-stage processes while allowing for shared inputs and assuming constant returns to scale. Liu et al. (2015) presented a two-stage DEA model with undesirable input-intermediate-outputs.

Mahdiloo et al. (2016) proposed a multiple criteria DEA approach for evaluating two-stage processes. The two-stage non-radial DEA model of Wang et al. (2016) can be used to define indices capable of measuring R&D efficiency, market efficiency, and integrated innovation efficiency. Izadikhah and Farzipoor Saen (2016) presented a new two-stage DEA model considering negative input-intermediate-output data. The DEA model presented by Wu et al. (2016) finds applications in environmental efficiency evaluations of two-stage systems where undesired outputs are allowed. Zhou, Lin, Xiao, Ma, and Wu (2017) presented an extension of the conventional DEA formulations by considering the serial relationship characterizing two-stage processes when implementing stochastic data. Fukuyama and Matousek (2017) developed a two-stage DEA network cost function model to evaluate banks' network revenue performance. Li and Cui (2017) proposed a network range adjusted environmental DEA to discuss environmental inefficiency changes.

2.2. Efficiency evaluation of banking systems using two-stage DEA

DEA is a systematic approach for analyzing the performance of organizations that was first proposed by Charnes, Cooper, and Rhodes (1978). In the traditional DEA methods, the internal structure of the DMUs is ignored. However, in many cases, the DMUs under evaluation consist of two-stage network structures with intermediate products.

The work of Charnes et al. (1986) on army recruitment using a two-stage approach was the first study to discuss the loss of information intrinsic to single-stage models. For what concerns banking efficiency evaluation models, the first two-stage DEA model was proposed by Seiford and Zhu (1999) who implemented their model to evaluate the marketability and profitability of a group of U.S. commercial banks. The next important work focusing on the evaluation problem of banking systems via two-stage DEA models is due to Zha and Liang (2010). Their method allows for shared inputs to be freely allocated among different stages and was applied to evaluate 30 U.S. commercial banks. Du et al. (2011) developed a Nash bargaining game model to measure the performance of commercial banks with a two-stage structure. Akther et al. (2013) defined a two-stage slacks-based inefficiency measure and used it in combination with the directional technology distance function to assess 19 private commercial banks and 2 government-owned banks in Bangladesh. Tavana and Khalili-Damghani (2014) proposed a two-stage fuzzy DEA model to calculate the efficiency scores of 20 bank branches. Wanke and Barros (2014) presented a two-stage process for measuring efficiency in the Brazilian banking industry.

Liu et al. (2015) proposed a two-stage DEA model with undesirable input-intermediate-outputs and evaluated 16 listed commercial banks in China. Huang, Lin, and Chen (2017) proposed a stochastic two-stage model within the stochastic frontier approach. They applied the model using data from the Chinese banking industry. Zhou et al. (2017) proposed a stochastic two-stage network DEA model based on a centralized control organization mechanism and measured the efficiencies of commercial banks in China. Degl'Innocenti, Kourtzidis, Sevic, and Tzeremes (2017) extended the Malmquist Productivity Index by applying an additive two-stage DEA model for assessing the bank productivity growth and the integration process of 28 countries of the European Union.

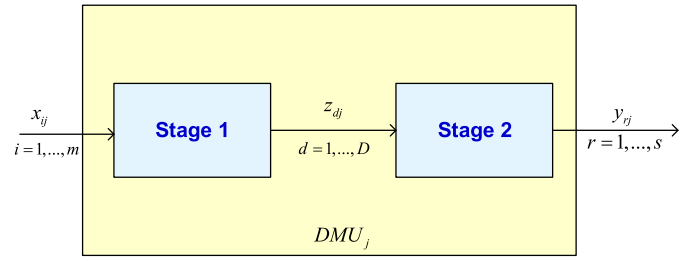

Fig. 1. A two-stage DMU.

The above-mentioned studies have proposed some novel ideas for improving two-stage DEA models in the banking industry. However, the literature on this topic is remarkably vast. Among the many studies focusing on the application of existing two-stage DEA models to the banking industry, it is worth citing Matthews (2013), Wang et al. (2014), Avkiran (2015), Ohsato and Takahashi (2015), Kwon and Lee (2015), Fukuyama and Matousek (2017), Ding, Dong, Liang, and Zhu (2017), and Wanke, Maredza, and Gupta (2017).

3. Methodological background

Consider n decision making units (DMUs) each of which uses m inputs to produce s outputs. A two-stage DMU is a DMU consisting of two sub-DMUs working in series. The outputs of the first stage (intermediate products) are usually used as inputs of the second stage to produce the final outputs. Finally, all inputs and outputs are assigned a weight. The following notations are used to review three keynote models in the literature that will be extended by our approach.

DMU_j	$j=1,\ 2,\ \ldots,\ n$	jth DMU
DMU_p		Fixed but generic DMU
x_{ij}	$i=1,\ 2,\ \ldots,\ m,$	ith input of Stage 1 (initial input) of DMU_j
z_{dj}	$d=1,\ 2,\ \ldots,\ D,$	dth output of Stage 1 (intermediate output) of DMU_j
y_{rj}	$r=1,\ 2,\ \ldots,\ s,$	rth output of Stage 2 (final output) of DMU_j
x^2_{hj}	$h=1,\ 2,\ \ldots,\ H,$	Additional hth input of Stage 2 of DMU_j
v_i	$i=1,\ 2,\ \ldots,\ m$	Weight of the ith initial input
w_d	$d=1,\ 2,\ \ldots,\ D$	Weight of the dth intermediate output
u_r	$r=1,\ 2,\ \ldots,\ s$	Weight of the rth final output
q_h	$h=1,\ 2,\ \ldots,\ H$	Weight of the hth additional input
θ_p		Overall efficiency of the generic pth DMU, DMU_p
θ_{1p}		Efficiency of Stage 1 (the first sub-DMU) of DMU_p
θ_{2p}		Efficiency of Stage 2 (the second sub-DMU) of DMU_p
θ^U_{1p}		Upper bound of the efficiency of Stage 1 of DMU_p
θ^U_{2p}		Upper bound of the efficiency of Stage 2 of DMU_p
θ^L_{1p}		Lower bound of the efficiency of Stage 1 of DMU_p
θ^L_{2p}		Lower bound of the efficiency of Stage 2 of DMU_p
ε		Non-Archimedean positive small number

Finally, the asterisk symbol (*) will be added as upper index to denote the optimal value of the corresponding efficiency score.

3.1. Kao and Hwang method

Kao and Hwang (2008) investigated the efficiency decomposition approach for a two-stage production process where the outputs of the first stage are the only inputs of the second stage. Their model takes into account the sequentiality of the whole process focusing on the relationships existing between its two sub-processes. Fig. 1 outlines the structure of a two-stage DMU of a production process.

In Kao and Hwang (2008) framework, each DMU consists of two sub-DMUs in series and its overall efficiency can be decomposed in

the product of the efficiencies of the two sub-DMUs. Thus, the following linear programming model is proposed for calculating the overall efficiency of the generic pth DMU, DMU_p:

$$\begin{aligned}
&\theta_p = \max \sum_{d=1}^{D} w_d z_{dp} \\
&s.t. \\
&\sum_{i=1}^{m} v_i x_{ip} = 1, \\
&\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \le 0, \qquad \forall j = 1, \ldots, n; \\
&\sum_{d=1}^{D} w_d z_{dj} - \sum_{i=1}^{m} v_i x_{ij} \le 0, \qquad \forall j = 1, \ldots, n; \\
&\sum_{r=1}^{s} u_r y_{rj} - \sum_{d=1}^{D} w_d z_{dj} \le 0, \qquad \forall j = 1, \ldots, n; \\
&u_r, v_i, w_d \ge \varepsilon, \qquad \forall r = 1, \ldots, s; i = 1, \ldots, m; d = 1, \ldots, D.
\end{aligned} \tag{1}$$

They introduced the non-Archimedean positive small number, ε, to prevent variables from taking a zero value and also to increase the discrimination power of the model. For a survey on the role of ε in DEA models, one can refer to Amin and Toloo (2004), Toloo (2014), and Toloo and Tavana (2017). The efficiency of the first sub-DMU (i.e., Stage 1) of DMU_p is obtained by another linear programming model:

$$\begin{aligned}
&\theta_{1p} = \max \sum_{d=1}^{D} w_d z_{dp} \\
&s.t. \\
&\sum_{i=1}^{m} v_i x_{ip} = 1, \\
&\sum_{r=1}^{s} u_r y_{rp} - \theta_p \sum_{i=1}^{m} v_i x_{ip} = 0, \\
&\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \le 0, \qquad \forall j = 1, \ldots, n; \\
&\sum_{d=1}^{D} w_d z_{dj} - \sum_{i=1}^{m} v_i x_{ij} \le 0, \qquad \forall j = 1, \ldots, n; \\
&\sum_{r=1}^{s} u_r y_{rj} - \sum_{d=1}^{D} w_d z_{dj} \le 0, \qquad \forall j = 1, \ldots, n; \\
&u_r, v_i, w_d \ge \varepsilon, \qquad \forall r = 1, \ldots, s; i = 1, \ldots, m; d = 1, \ldots, D.
\end{aligned} \tag{2}$$

Finally, the efficiency of the second sub-DMU (i.e., Stage 2) is obtained as: $\theta_{2p} = \theta_p / \theta_{1p}$.

3.2. Zha and Liang method

Zha and Liang (2010) designed a method to model shared flow in two-stage production processes, where the inputs are assumed to be freely allocated in either one of the two stages linked by a cooperative relationship. To model this kind of processes, an allocative factor α_i is assigned to each input quantity $x_{ij}$of a given DMU_j. The factor α_i divides the input quantity x_{ij} in two parts, $\alpha_i x_{ij}$ and $(1-\alpha_i)x_{ij}$, to be used as input in Stage 1 and Stage 2, respectively. Fig. 2 outlines the structure of a two-stage DMU with shared inputs.

Zha and Liang (2010) proposed the following product-form cooperative efficiency model to compute the overall efficiency of the generic p-th DMU, DMU_p, and express the relationships between

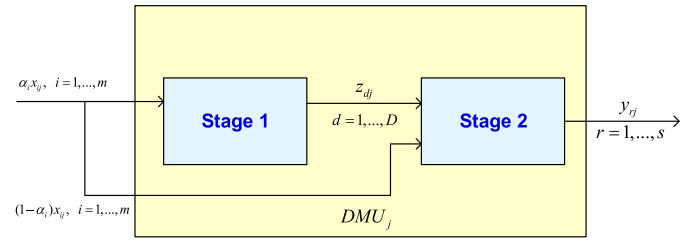

Fig. 2. A two-stage DMU with shared inputs.

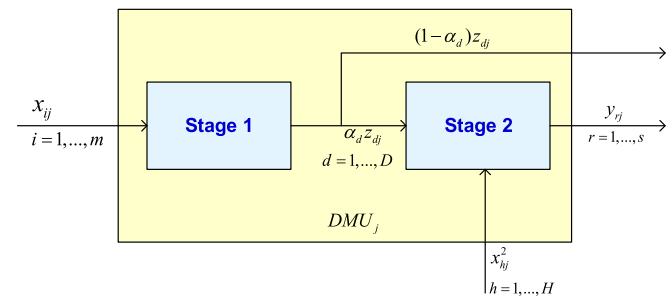

Fig. 3. A two-stage DMU with shared intermediate outputs and extra inputs in Stage 2.

the two stages of the production process.

$$\begin{aligned}
&\theta_p = \max \sum_{d=1}^{D} w_d z_{dp} \times \sum_{r=1}^{s} u_r y_{rp} \\
&s.t. \\
&\sum_{i=1}^{m} \pi_i x_{ip} = 1, \\
&\delta \left[\sum_{i=1}^{m} (v_i - \pi_i) x_{ip} + \sum_{d=1}^{D} w_d z_{dp} \right] = 1, \\
&\sum_{r=1}^{s} u_r y_{rj} - \delta \left[\sum_{i=1}^{m} (v_i - \pi_i) x_{ij} + \sum_{d=1}^{D} w_d z_{dj} \right] \le 0, \qquad \forall j = 1, \ldots, n; \\
&\sum_{d=1}^{D} w_d z_{dj} - \sum_{i=1}^{m} \pi_i x_{ij} \le 0, \qquad \forall j = 1, \ldots, n; \\
&v_i \ge \pi_i \ge 0, \qquad \forall i = 1, \ldots, m. \\
&u_r, w_d \ge \varepsilon, \qquad \forall r = 1, \ldots, s; d = 1, \ldots, D.
\end{aligned} \tag{3}$$

where $\pi_i = v_i \alpha_i$ $(i = 1, 2, \ldots, m)$ is the weight of the portion of the ith input used in Stage 1.

Then, they applied a heuristic method to transform the above nonlinear programming model into a parametric linear one, where δ is considered to be a parameter, which can be used to solve the cooperative problem.

3.3. Yu and Shi method

Yu and Shi (2014) proposed a two-stage DEA model with shared intermediate products and additional inputs in the second stage, where the outputs of the first stage are used both as inputs of the second stage and as final outputs. Fig. 3 describes a two-stage DMU whose intermediate products by the sub-DMU in Stage 1 are partially consumed by the sub-DMU in Stage 2 (together with additional inputs) and partially used as final outputs.

Similarly to the case of shared inputs, an allocative factor α_d is assumed for each intermediate output quantity z_{dj} of a given DMU_j and used to split the output z_{dj} in two parts, $\alpha_d z_{dj}$ and $(1-\alpha_d)z_{dj}$,

which become inputs of Stage 2 and final outputs, respectively. Moreover, $\varphi_d = w_d\alpha_d$ $(d=1,\ 2,\ ...,\ D)$ denotes the weight of the portion of the dth intermediate output used in Stage 2.

In order to determine the upper and lower bounds of the efficiencies of the two sub-DMUs composing a unit, Yu and Shi (2014) adopted a non-cooperative approach by considering two distinct leader-follower cases.

Case 1: *Sub-DMU in Stage 1 dominates the system, while the sub-DMU in Stage 2 follows*

In this case, the sub-DMU in Stage 1 is assumed to be the leader. That is, the performance of Stage 1 is more important than that of Stage 2 (follower) and the efficiency of the sub-DMU in Stage 2 is computed subject to the condition that the Stage 1 (leader) efficiency remains constant.

The upper bound of the efficiency of Stage 1 and the lower bound of the efficiency of Stage 2 of the generic p-th DMU, DMU_p, were modelled using two fractional programs that were in turn converted into the following linear programming problems by applying the Charnes–Cooper transformations (Charnes & Cooper, 1962).

$$\begin{aligned}
&\theta_{1p}^{U} = \max \sum_{d=1}^{D} w_d z_{dp} \\
&s.t. \\
&\sum_{i=1}^{m} v_i x_{ip} = 1, \\
&\sum_{d=1}^{D} w_d z_{dj} - \sum_{i=1}^{m} v_i x_{ij} \le 0, \qquad \forall j = 1, ..., n; \\
&w_d, v_i \ge 0, \qquad \forall d = 1, ..., D; i = 1, ...m.
\end{aligned} \tag{4}$$

and

$$\begin{aligned}
&\theta_{2p}^{L} = \max \sum_{r=1}^{s} u_r y_{rp} \\
&s.t. \\
&\sum_{d=1}^{D} \varphi_d z_{dp} + \sum_{h=1}^{H} q_h x_{hp}^2 = 1, \\
&\sum_{r=1}^{s} u_r y_{rj} - \sum_{d=1}^{D} \varphi_d z_{dj} - \sum_{h=1}^{H} q_h x_{hj}^2 \le 0, \qquad \forall j = 1, ..., n; \\
&\sum_{d=1}^{D} w_d z_{dp} - \theta_{1p}^{U*} \times \sum_{i=1}^{m} v_i x_{ip} = 0, \\
&\sum_{d=1}^{D} w_d z_{dj} - \sum_{i=1}^{m} v_i x_{ij} \le 0, \qquad \forall j = 1, ..., n; \\
&w_d \ge \varphi_d \ge 0, \forall d = 1, ..., D. \\
&u_r, \varphi_d, q_h, w_d, v_i \ge 0, \forall r{=}1, .., s; d{=}1, .., D; h{=}1, .., H; i{=}1, .., m.
\end{aligned} \tag{5}$$

Case 2: *Sub-DMU in Stage 2 dominates the system, while the sub-DMU in Stage 1 follows*

In this case, the sub-DMU in Stage 2 is assumed to be the leader. Hence, the efficiency of the sub-DMU in Stage 1 must be computed subject to the condition that the Stage 2 (leader) efficiency remains constant.

Two fractional programs were proposed to calculate the upper bound of the efficiency of Stage 2 and the lower bound of the efficiency of Stage 1. Again, these problems were converted into the following linear programming problems using the Charnes–Cooper transformations.

$$\begin{aligned}
&\theta_{2p}^{U} = \max \sum_{r=1}^{s} u_r y_{rp} \\
&s.t. \\
&\sum_{d=1}^{D} \varphi_d z_{dp} + \sum_{h=1}^{H} q_h x_{hp}^2 = 1, \\
&\sum_{r=1}^{s} u_r y_{rj} - \sum_{d=1}^{D} \varphi_d z_{dj} - \sum_{h=1}^{H} q_h x_{hj}^2 \le 0, \qquad \forall j = 1, ..., n; \\
&u_r, \varphi_d, q_h \ge 0, \quad \forall r = 1, .., s; d = 1, .., D; h = 1, .., H.
\end{aligned} \tag{6}$$

and

$$\begin{aligned}
&\theta_{1p}^{L} = \max \sum_{d=1}^{D} \varphi_d z_{dp} \\
&s.t. \\
&\sum_{i=1}^{m} v_i x_{ip} = 1, \\
&\sum_{d=1}^{D} \varphi_d z_{dp} + \sum_{h=1}^{H} q_h x_{hp}^2 = 1, \\
&\sum_{r=1}^{s} u_r y_{rp} = \theta_{2p}^{U*}, \\
&\sum_{d=1}^{D} w_d z_{dj} \le \sum_{i=1}^{m} v_i x_{ij}, \qquad \forall j = 1, ..., n; \\
&\sum_{r=1}^{s} u_r y_{rj} \le \sum_{d=1}^{D} \varphi_d z_{dj} + \sum_{h=1}^{H} q_h x_{hj}^2, \\
&w_d \ge \varphi_d \ge 0, \forall d = 1, ..., D. \\
&u_r, \varphi_d, q_h, w_d, v_i \ge 0, \forall r{=}1, .., s; d{=}1, .., D; h{=}1, .., H; i = 1, .., m.
\end{aligned} \tag{7}$$

Hence, Yu and Shi (2014) focused on the cooperative efficiency of the generic pth DMU, DMU_p. A nonlinear fractional programming problem was used to model the overall efficiency of DMU_p and the Charnes–Cooper transformation applied to restate it as follows:

$$\begin{aligned}
&\theta_p = \max \sum_{d=1}^{D} \varphi_d z_{dp} \times \sum_{r=1}^{s} u_r y_{rp} \\
&s.t. \\
&\sum_{i=1}^{m} v_i x_{ip} = 1, \\
&\sum_{d=1}^{D} \varphi_d z_{dp} + \sum_{h=1}^{H} q_h x_{hp}^2 = 1, \\
&\sum_{d=1}^{D} w_d z_{dj} \le \sum_{i=1}^{m} v_i x_{ij}, \qquad \forall j = 1, ..., n; \\
&\sum_{r=1}^{s} u_r y_{rj} \le \sum_{d=1}^{D} \varphi_d z_{dj} + \sum_{h=1}^{H} q_h x_{hj}^2, \qquad \forall j = 1, ..., n; \\
&w_d \ge \varphi_d \ge 0, \qquad \forall d = 1, ..., D. \\
&u_r, \varphi_d, q_h, w_d, v_i \ge 0, \quad \forall r{=}1, .., s; d{=}1, .., D; h{=}1, .., H; i{=}1, .., m.
\end{aligned} \tag{8}$$

Model (8) is again a nonlinear programming problem. Thus, Yu and Shi (2014) set $\lambda = \sum_{d=1}^{D} \varphi_d z_{dp}$ and converted it into the fol-

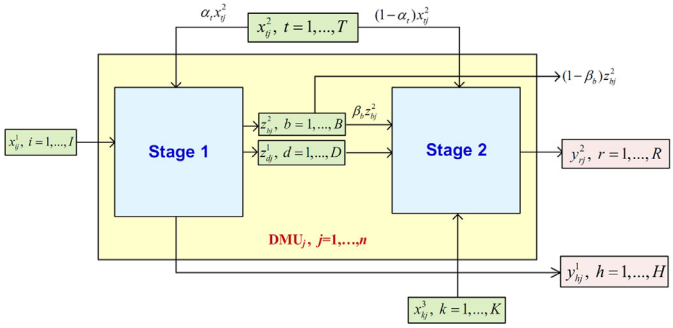

Fig. 4. The proposed two-stage production process.

lowing parametric linear programming problem:

$$\theta_p = \max \ \lambda \times \sum_{r=1}^{s} u_r y_{rp}$$

$$s.t.$$

$$\sum_{i=1}^{m} v_i x_{ip} = 1,$$

$$\sum_{d=1}^{D} \varphi_d z_{dp} + \sum_{h=1}^{H} q_h x_{hp}^2 = 1,$$

$$\sum_{d=1}^{D} w_d z_{dj} \le \sum_{i=1}^{m} v_i x_{ij}, \qquad \forall j = 1, \ldots, n;$$

$$\sum_{r=1}^{s} u_r y_{rj} \le \sum_{d=1}^{D} \varphi_d z_{dj} + \sum_{h=1}^{H} q_h x_{hj}^2, \qquad \forall j = 1, \ldots, n;$$

$$w_d \ge \varphi_d \ge 0, \qquad \forall d = 1, \ldots, D.$$

$$u_r, \varphi_d, q_h, w_d, v_i \ge 0, \quad \forall r=1, .., s; d=1, \ldots, D; h=1, \ldots, H; i = 1, \ldots, m. \tag{9}$$

where $\lambda \in [\theta_{1p}^L, \theta_{1p}^U]$ can be treated as a parameter. An alternative transformation of Model (8) can be carried out by setting $\lambda = \sum_{r=1}^{s} u_r y_{rp}$ and $\theta_{2p}^L \le \lambda \le \theta_{2p}^U$.

4. Proposed methodology

We consider a two-stage DEA model where each DMU consists of two sub-DMUs in series with the following characteristics:

- intermediate outputs of the sub-DMU in Stage 1 are only partially consumed by the sub-DMU in Stage 2;
- part of the inputs of the DMU can be freely allocated to each one of the two sub-DMUs;
- additional inputs are directly consumed in Stage 2;
- part of the outputs of Stage 1 are considered as final outputs.

Fig. 4 depicts a two-stage DMU such as the one just described.

In summary, there are three categories of inputs, two categories of intermediate outputs and two categories of final outputs:

$x_{ij}^1, i = 1, \ldots, I$ Inputs consumed by Stage 1 of DMU_j entirely.
$x_{tj}^2, t = 1, \ldots, T$ Inputs freely allocated to both sub-DMUs of DMU_j.
$x_{kj}^3, k = 1, \ldots, K$ Inputs consumed by Stage 2 of DMU_j entirely.
$z_{dj}^1, d = 1, \ldots, D$ Intermediate outputs consumed by Stage 2 of DMU_j entirely.
$z_{bj}^2, b = 1, \ldots, B$ Intermediate outputs acting as inputs of Stage 2 and final outputs of DMU_j.
$y_{hj}^1, h = 1, \ldots, H$ Final outputs produced by Stage 1 of DMU_j.
$y_{rj}^2, r = 1, \ldots, R$ Final outputs produced by Stage 2 of DMU_j.

The weights of the inputs x_{ij}^1, x_{tj}^2 and x_{kj}^3 will be denoted by v_{ij}^1, v_{tj}^2 and v_{kj}^3, respectively. Similarly, the weights of the intermediate outputs z_{dj}^1 and z_{bj}^2 will be denoted by w_{dj}^1 and w_{bj}^2, while u_{hj}^1 and u_{rj}^2 will stand for the weights of the final outputs y_{hj}^1 and y_{rj}^2, respectively,

Also, two sets of allocative factors are assumed in order to split the shared inputs and the shared intermediate outputs:

α_t, $t = 1, \ldots, T$

Allocative factors dividing x_{tj}^2 into two parts:

- $\alpha_t x_{tj}^2$ is the input of the sub-DMU in Stage 1,
- $(1 - \alpha_t) x_{tj}^2$ is the input of the sub-DMU in Stage 2.

β_b, $b = 1, \ldots, B$

Allocative factors dividing z_{bj}^2 into two parts:

- $\beta_b z_{bj}^2$ is the input of the sub-DMU in Stage 2,
- $(1 - \beta_b) z_{bj}^2$ is the final output.

4.1. The non-cooperative model

In this section, we build on the Stackelberg game and the concept of leader–follower (Liang et al., 2006, 2008; Yu and Shi (2014)) to develop a performance measuring model. We start by determining the upper and lower bounds of the efficiencies of the sub-DMUs in the different stages of the non-cooperative setting. Following Yu and Shi (2014), we analyze two cases.

4.1.1. Sub-DMU in Stage 1 dominates the system, while the sub-DMU in Stage 2 follows

The leader's efficiency:

To evaluate the efficiency of the sub-DMU in Stage 1, we propose Model (10) below. Model (10) allows to determine the upper bound of the efficiency of the sub-DMU in Stage 1 for the generic DMU under evaluation, DMU_p.

$$\theta_{1p}^U = \max \ \frac{\sum_{d=1}^{D} w_d^1 z_{dp}^1 + \sum_{b=1}^{B} w_b^2 z_{bp}^2 + \sum_{h=1}^{H} u_h^1 y_{hp}^1}{\sum_{i=1}^{I} v_i^1 x_{ip}^1 + \sum_{t=1}^{T} v_t^2 (\alpha_t x_{tp}^2)}$$

$$s.t.$$

$$\frac{\sum_{d=1}^{D} w_d^1 z_{dj}^1 + \sum_{b=1}^{B} w_b^2 z_{bj}^2 + \sum_{h=1}^{H} u_h^1 y_{hj}^1}{\sum_{i=1}^{I} v_i^1 x_{ij}^1 + \sum_{t=1}^{T} v_t^2 (\alpha_t x_{tj}^2)} \le 1, \qquad \forall j = 1, \ldots, n,$$

$$w_d^1, w_b^2, u_h^1, v_i^1, v_t^2 \ge 0, \qquad \forall d = 1, .., D; b = 1, .., B; \ h = 1, .., H; i = 1, .., m; t = 1, .., T.$$

$$0 \le \alpha_t \le 1, \qquad \forall t = 1, \ldots, T. \tag{10}$$

By using the Charnes–Cooper transformation Charnes & Cooper, 1962) and setting $\mu_t^2 = v_t^2 \alpha_t$ ($t = 1, \ldots, T$), Model ((10) can be transformed into the following linear model:

$$\theta_{1p}^{U} = \max \quad \sum_{d=1}^{D} w_d^1 z_{dp}^1 + \sum_{b=1}^{B} w_b^2 z_{bp}^2 + \sum_{h=1}^{H} u_h^1 y_{hp}^1$$

$$s.t.$$

$$\sum_{i=1}^{I} v_i^1 x_{ip}^1 + \sum_{t=1}^{T} \mu_t^2 x_{tp}^2 = 1,$$

$$\sum_{d=1}^{D} w_d^1 z_{dj}^1 + \sum_{b=1}^{B} w_b^2 z_{bj}^2 + \sum_{h=1}^{H} u_h^1 y_{hj}^1 - \sum_{i=1}^{I} v_i^1 x_{ij}^1 - \sum_{t=1}^{T} \mu_t^2 x_{tj}^2 \le 0, \quad \forall j = 1, ..., n,$$

$$w_d^1, w_b^2, u_h^1, v_i^1, \mu_t^2 \ge 0, \quad \forall d = 1, .., D; b = 1, .., B; h = 1, .., H; i = 1, .., m; t = 1, .., T. \tag{11}$$

The follower's efficiency:

To evaluate the efficiency of the sub-DMU in Stage 2, we propose Model (12) below. Solving Model (12) yields the lower bound of the efficiency of the sub-DMU in Stage 2 for the generic DMU under evaluation, DMU_p.

$$\theta_{2p}^{L} = \max \quad \frac{\sum_{r=1}^{R} u_r^2 y_{rp}^2}{\sum_{k=1}^{K} v_k^3 x_{kp}^3 + \sum_{t=1}^{T} v_t^2 (1-\alpha_t) x_{tp}^2 + \sum_{d=1}^{D} w_d^1 z_{dp}^1 + \sum_{b=1}^{B} w_b^2 (\beta_b z_{bp}^2)}$$

$$s.t.$$

$$\frac{\sum_{r=1}^{R} u_r^2 y_{rj}^2}{\sum_{k=1}^{K} v_k^3 x_{kj}^3 + \sum_{t=1}^{T} v_t^2 (1-\alpha_t) x_{tj}^2 + \sum_{d=1}^{D} w_d^1 z_{dj}^1 + \sum_{b=1}^{B} w_b^2 (\beta_b z_{bj}^2)} \le 1, \quad \forall j = 1, ..., n,$$

$$\frac{\sum_{d=1}^{D} w_d^1 z_{dp}^1 + \sum_{b=1}^{B} w_b^2 z_{bp}^2 + \sum_{h=1}^{H} u_h^1 y_{hp}^1}{\sum_{i=1}^{I} v_i^1 x_{ip}^1 + \sum_{t=1}^{T} v_t^2 (\alpha_t x_{tp}^2)} = \theta_{1p}^{U*},$$

$$\frac{\sum_{d=1}^{D} w_d^1 z_{dj}^1 + \sum_{b=1}^{B} w_b^2 z_{bj}^2 + \sum_{h=1}^{H} u_h^1 y_{hj}^1}{\sum_{i=1}^{I} v_i^1 x_{ij}^1 + \sum_{t=1}^{T} v_t^2 (\alpha_t x_{tj}^2)} \le 1, \quad \forall j = 1, .., n,$$

$$w_d^1, w_b^2, u_r^2, v_k^3, v_t^2, u_h^1, v_i^1 \ge 0, \quad \forall d = 1, .., D; b = 1, .., B; r = 1, .., s; k = 1, .., K; h = 1, .., H; i = 1, .., I; t = 1, .., T.$$

$$0 \le \alpha_t \le 1, \quad \forall t = 1, ..., T,$$

$$0 \le \beta_b \le 1, \quad \forall b = 1, ..., B. \tag{12}$$

By using the Charnes–Cooper transformation (Charnes & Cooper, 1962) and setting $\mu_t^2 = v_t^2 \alpha_t$ ($t = 1, ..., T$) and $\varphi_b^2 = w_b^2 \beta_b$ ($b = 1, ..., B$), Model (12) can be transformed into the following linear model:

$$\theta_{2p}^{L} = \max \quad \sum_{r=1}^{R} u_r^2 y_{rp}^2$$

$$s.t.$$

$$\sum_{k=1}^{K} v_k^3 x_{kp}^3 + \sum_{t=1}^{T} (v_t^2 - \mu_t^2) x_{tp}^2 + \sum_{d=1}^{D} w_d^1 z_{dp}^1 + \sum_{b=1}^{B} \varphi_b^2 z_{bp}^2 = 1,$$

$$\sum_{r=1}^{R} u_r^2 y_{rj}^2 - \sum_{k=1}^{K} v_k^3 x_{kj}^3 - \sum_{t=1}^{T} (v_t^2 - \mu_t^2) x_{tj}^2 - \sum_{d=1}^{D} w_d^1 z_{dj}^1 - \sum_{b=1}^{B} \varphi_b^2 z_{bj}^2 \le 0, \quad \forall j = 1, ..., n,$$

$$\sum_{d=1}^{D} w_d^1 z_{dp}^1 + \sum_{b=1}^{B} w_b^2 z_{bp}^2 + \sum_{h=1}^{H} u_h^1 y_{hp}^1 - \theta_{1p}^{U*} \left(\sum_{i=1}^{I} v_i^1 x_{ip}^1 + \sum_{t=1}^{T} \mu_t^2 x_{tp}^2\right) = 0,$$

$$\sum_{d=1}^{D} w_d^1 z_{dj}^1 + \sum_{b=1}^{B} w_b^2 z_{bj}^2 + \sum_{h=1}^{H} u_h^1 y_{hj}^1 - \sum_{i=1}^{I} v_i^1 x_{ij}^1 - \sum_{t=1}^{T} \mu_t^2 x_{tj}^2 \le 0, \quad \forall j = 1, ..., n,$$

$$w_d^1, w_b^2, u_r^2, v_k^3, v_t^2, u_h^1, v_i^1, \mu_t^2, \varphi_b^2 \ge 0, \quad \forall d = 1, .., D; b = 1, .., B; r = 1, .., s; k = 1, .., K; h = 1, .., H; i = 1, .., I; t = 1, .., T.$$

$$0 \le \mu_t^2 \le v_t^2, \quad \forall t = 1, ..., T,$$

$$0 \le \varphi_b^2 \le w_b^2, \quad \forall b = 1, ..., B. \tag{13}$$

4.1.2. Sub-DMU in Stage 2 dominates the system, while the sub-DMU in Stage 1 follows

The leader's efficiency:

To evaluate the efficiency of the sub-DMU in Stage 2, we propose Model (14) below. Model (14) allows to determine the upper bound of the efficiency of the sub-DMU in Stage 2 for the generic DMU under evaluation, DMU_p.

$$\theta_{2p}^{U} = \max \quad \frac{\sum_{r=1}^{R} u_r^2 y_{rp}^2}{\sum_{k=1}^{K} v_k^3 x_{kp}^3 + \sum_{t=1}^{T} v_t^2 (1-\alpha_t) x_{tp}^2 + \sum_{d=1}^{D} w_d^1 z_{dp}^1 + \sum_{b=1}^{B} w_b^2 (\beta_b z_{bp}^2)}$$

$$s.t.$$

$$\frac{\sum_{r=1}^{R} u_r^2 y_{rj}^2}{\sum_{k=1}^{K} v_k^3 x_{kj}^3 + \sum_{t=1}^{T} v_t^2 (1-\alpha_t) x_{tj}^2 + \sum_{d=1}^{D} w_d^1 z_{dj}^1 + \sum_{b=1}^{B} w_b^2 (\beta_b z_{bj}^2)} \le 1, \quad \forall j = 1, ..., n,$$

$$w_d^1, w_b^2, u_r^2, v_k^3, v_t^2 \ge 0, \quad \forall d = 1, .., D; b = 1, .., B; r = 1, .., s; t = 1, .., T.$$

$$0 \le \alpha_t \le 1, \quad \forall t = 1, ..., T,$$

$$0 \le \beta_b \le 1, \quad \forall b = 1, ..., B. \tag{14}$$

By using the Charnes–Cooper transformation and setting $\pi_t^2 = v_t^2 (1-\alpha_t)$ and $\varphi_b^2 = w_b^2 \beta_b$, Model (14) can be transformed into the following linear model:

$$\theta_{2p}^{U} = \max \quad \sum_{r=1}^{R} u_r^2 y_{rp}^2$$

$$s.t.$$

$$\sum_{k=1}^{K} v_k^3 x_{kp}^3 + \sum_{t=1}^{T} \pi_t^2 x_{tp}^2 + \sum_{d=1}^{D} w_d^1 z_{dp}^1 + \sum_{b=1}^{B} \varphi_b^2 z_{bp}^2 = 1,$$

$$\sum_{r=1}^{R} u_r^2 y_{rj}^2 - \sum_{k=1}^{K} v_k^3 x_{kj}^3 - \sum_{t=1}^{T} \pi_t^2 x_{tj}^2 - \sum_{d=1}^{D} w_d^1 z_{dj}^1 - \sum_{b=1}^{B} \varphi_b^2 z_{bj}^2 \le 0, \quad \forall j = 1, ..., n,$$

$$w_d^1, u_r^2, v_k^3, \varphi_b^2, \pi_t^2 \ge 0, \quad \forall d = 1, .., D; b = 1, .., B; r = 1, .., s; k = 1, ..K; t = 1, .., T. \tag{15}$$

The follower's efficiency:

Model (16) evaluates the efficiency of the sub-DMU in Stage 1 interpreted as the follower of Stage 2, that is, the lower bound of

the efficiency of the sub-DMU in Stage 1 for the generic DMU under evaluation, DMU_p.

$$\theta_{1p}^L = \max \quad \frac{\sum_{d=1}^{D} w_d^1 z_{dp}^1 + \sum_{b=1}^{B} w_b^2 z_{bp}^2 + \sum_{h=1}^{H} u_h^1 y_{hp}^1}{\sum_{i=1}^{I} v_i^1 x_{ip}^1 + \sum_{t=1}^{T} v_t^2 (\alpha_t x_{tp}^2)}$$

$$s.t.$$

$$\frac{\sum_{d=1}^{D} w_d^1 z_{dj}^1 + \sum_{b=1}^{B} w_b^2 z_{bj}^2 + \sum_{h=1}^{H} u_h^1 y_{hj}^1}{\sum_{i=1}^{I} v_i^1 x_{ij}^1 + \sum_{t=1}^{T} v_t^2 (\alpha_t x_{tj}^2)} \le 1, \qquad \forall j = 1, ..., n,$$

$$\frac{\sum_{r=1}^{R} u_r^2 y_{rp}^2}{\sum_{k=1}^{K} v_k^3 x_{kp}^3 + \sum_{t=1}^{T} v_t^2 (1-\alpha_t) x_{tp}^2 + \sum_{d=1}^{D} w_d^1 z_{dp}^1 + \sum_{b=1}^{B} w_b^2 (\beta_b z_{bp}^2)} = \theta_{2p}^{U*},$$

$$\frac{\sum_{r=1}^{R} u_r^2 y_{rj}^2}{\sum_{k=1}^{K} v_k^3 x_{kj}^3 + \sum_{t=1}^{T} v_t^2 (1-\alpha_t) x_{tj}^2 + \sum_{d=1}^{D} w_d^1 z_{dj}^1 + \sum_{b=1}^{B} w_b^2 (\beta_b z_{bj}^2)} \le 1, \qquad \forall j = 1, ..., n,$$

$$w_d^1, w_b^2, u_r^2, v_k^3, v_t^2, v_i^1, u_h^1 \ge 0, \quad \forall d = 1, .., D; b = 1, .., B; r = 1, .., s; k = 1, ..K; t = 1, .., T; i = 1, .., I; h = 1, .., H.$$

$$0 \le \alpha_t \le 1, \qquad \forall t = 1, ..., T,$$

$$0 \le \beta_b \le 1, \qquad \forall b = 1, ..., B. \tag{16}$$

Model (16) is a fractional programming problem. Therefore, we can use again the Charnes–Cooper transformation to convert it into a linear one. We also set $\mu_t^2 = v_t^2 \alpha_t$ and $\varphi_b^2 = w_b^2 \beta_b$.

$$\theta_{1p}^L = \max \quad \sum_{d=1}^{D} w_d^1 z_{dp}^1 + \sum_{b=1}^{B} w_b^2 z_{bp}^2 + \sum_{h=1}^{H} u_h^1 y_{hp}^1$$

$$s.t.$$

$$\sum_{i=1}^{I} v_i^1 x_{ip}^1 + \sum_{t=1}^{T} \mu_t^2 x_{tp}^2 = 1,$$

$$\sum_{d=1}^{D} w_d^1 z_{dj}^1 + \sum_{b=1}^{B} w_b^2 z_{bj}^2 + \sum_{h=1}^{H} u_h^1 y_{hj}^1 - \sum_{i=1}^{I} v_i^1 x_{ij}^1 - \sum_{t=1}^{T} \mu_t^2 x_{tj}^2 \le 0, \qquad \forall j = 1, ..., n,$$

$$\sum_{r=1}^{R} u_r^2 y_{rp}^2 - \theta_{2p}^{U*} \Big(\sum_{k=1}^{K} v_k^3 x_{kp}^3 + \sum_{t=1}^{T} (v_t^2 - \mu_t^2) x_{tp}^2 + \sum_{d=1}^{D} w_d^1 z_{dp}^1 + \sum_{b=1}^{B} \varphi_b^2 z_{bp}^2 \Big) = 0,$$

$$\sum_{r=1}^{R} u_r^2 y_{rj}^2 - \sum_{k=1}^{K} v_k^3 x_{kj}^3 - \sum_{t=1}^{T} (v_t^2 - \mu_t^2) x_{tj}^2 - \sum_{d=1}^{D} w_d^1 z_{dj}^1 - \sum_{b=1}^{B} \varphi_b^2 z_{bj}^2 \le 0, \qquad \forall j = 1, ..., n,$$

$$w_d^1, w_b^2, u_r^2, v_k^3, v_t^2, v_i^1, u_h^1 \ge 0, \quad \forall d = 1, .., D; b = 1, .., B; r = 1, .., s; k = 1, ..K; t = 1, .., T; i = 1, .., I; h = 1, .., H.$$

$$0 \le \mu_t^2 \le v_t^2, \qquad \forall t = 1, ..., T,$$

$$0 \le \varphi_b^2 \le w_b^2, \qquad \forall b = 1, ..., B. \tag{17}$$

4.2. The cooperative model

We can now evaluate the cooperative efficiency of a two-stage production process where each DMU consists of two sub-DMUs in series with the characteristics described in Fig. 4. Building on the models presented in the previous sections (Section 4.1) and the decomposition approach (Sections 3.1 and 3.2), the following fractional programming model is proposed to determine the overall efficiency θ_p of the generic p-th DMU, DMU_p.

$$\theta_p = \max \frac{\sum_{d=1}^{D} w_d^1 z_{dp}^1 + \sum_{b=1}^{B} w_b^2 z_{bp}^2 + \sum_{h=1}^{H} u_h^1 y_{hp}^1}{\sum_{i=1}^{I} v_i^1 x_{ip}^1 + \sum_{t=1}^{T} v_t^2 (\alpha_t x_{tp}^2)} \times \frac{\sum_{r=1}^{R} u_r^2 y_{rp}^2}{\sum_{k=1}^{K} v_k^3 x_{kp}^3 + \sum_{t=1}^{T} v_t^2 (1-\alpha_t) x_{tp}^2 + \sum_{d=1}^{D} w_d^1 z_{dp}^1 + \sum_{b=1}^{B} w_b^2 (\beta_b z_{bp}^2)}$$

$$s.t.$$

$$\frac{\sum_{d=1}^{D} w_d^1 z_{dj}^1 + \sum_{b=1}^{B} w_b^2 z_{bj}^2 + \sum_{h=1}^{H} u_h^1 y_{hj}^1}{\sum_{i=1}^{I} v_i^1 x_{ij}^1 + \sum_{t=1}^{T} v_t^2 (\alpha_t x_{tj}^2)} \le 1, \qquad \forall j = 1, ..., n,$$

$$\frac{\sum_{r=1}^{R} u_r^2 y_{rj}^2}{\sum_{k=1}^{K} v_k^3 x_{kj}^3 + \sum_{t=1}^{T} v_t^2 (1-\alpha_t) x_{tj}^2 + \sum_{d=1}^{D} w_d^1 z_{dj}^1 + \sum_{b=1}^{B} w_b^2 (\beta_b z_{bj}^2)} \le 1, \qquad \forall j = 1, ..., n,$$

$$w_d^1, w_b^2, u_r^2, v_k^3, v_t^2, v_i^1, u_h^1 \ge 0, \quad \forall d = 1, .., D; b = 1, .., B; r = 1, .., s; k = 1, ..K; t = 1, .., T; i = 1, .., I; h = 1, .., H$$

$$0 \le \alpha_t \le 1, \qquad \forall t = 1, ..., T,$$

$$0 \le \beta_b \le 1, \qquad \forall b = 1, ..., B. \tag{18}$$

By applying the Charnes–Cooper transformation and setting $\mu_t^2 = v_t^2 \alpha_t$ and $\varphi_b^2 = w_b^2 \beta_b$, Model (18) can be converted to the following model.

$$\theta_p = \max \left(\sum_{d=1}^{D} w_d^1 z_{dp}^1 + \sum_{b=1}^{B} w_b^2 z_{bp}^2 + \sum_{h=1}^{H} u_h^1 y_{hp}^1 \right) \times \left(\sum_{r=1}^{R} u_r^2 y_{rp}^2 \right)$$

$$s.t.$$

$$\sum_{i=1}^{I} v_i^1 x_{ip}^1 + \sum_{t=1}^{T} \mu_t^2 x_{tp}^2 = 1,$$

$$\sum_{k=1}^{K} v_k^3 x_{kp}^3 + \sum_{t=1}^{T} (v_t^2 - \mu_t^2) x_{tp}^2 + \sum_{d=1}^{D} w_d^1 z_{dp}^1 + \sum_{b=1}^{B} \varphi_b^2 z_{bp}^2 = 1,$$

$$\sum_{d=1}^{D} w_d^1 z_{dj}^1 + \sum_{b=1}^{B} w_b^2 z_{bj}^2 + \sum_{h=1}^{H} u_h^1 y_{hj}^1 - \sum_{i=1}^{I} v_i^1 x_{ij}^1 - \sum_{t=1}^{T} \mu_t^2 x_{tj}^2 \le 0, \qquad \forall j = 1, ..., n,$$

$$\sum_{r=1}^{R} u_r^2 y_{rj}^2 - \sum_{k=1}^{K} v_k^3 x_{kj}^3 - \sum_{t=1}^{T} (v_t^2 - \mu_t^2) x_{tj}^2 - \sum_{d=1}^{D} w_d^1 z_{dj}^1 - \sum_{b=1}^{B} \varphi_b^2 z_{bj}^2 \le 0, \qquad \forall j = 1, ..., n,$$

$$w_d^1, w_b^2, u_r^2, v_k^3, v_t^2, v_i^1, u_h^1 \ge 0, \quad \forall d = 1, .., D; b = 1, .., B; r = 1, .., s;$$

$$k = 1, ..K; t = 1, .., T; i = 1, .., I; h = 1, .., H.$$
$$0 \le \varphi_b^2 \le w_b^2, \qquad \forall b = 1, ..., B,$$
$$0 \le \mu_t^2 \le v_t^2, \qquad \forall t = 1, ..., T. \tag{19}$$

Model (19) is a nonlinear programming problem and we have:

$$\theta_{1p}^L \le \sum_{d=1}^{D} w_d^1 z_{dp}^1 + \sum_{b=1}^{B} w_b^2 z_{bp}^2 + \sum_{h=1}^{H} u_h^1 y_{hp}^1 \le \theta_{1p}^U \text{ and}$$
$$\theta_{2p}^L \le \sum_{r=1}^{R} u_r^2 y_{rp}^2 \le \theta_{2p}^U.$$

Therefore, letting $\lambda = \sum_{d=1}^{D} w_d^1 z_{dp}^1 + \sum_{b=1}^{B} w_b^2 z_{bp}^2 + \sum_{h=1}^{H} u_h^1 y_{hp}^1$, Model (19) becomes the following nonlinear programming model:

$$\theta_p = \max \quad \lambda \times \left(\sum_{r=1}^{R} u_r^2 y_{rp}^2 \right)$$
$$s.t.$$
$$\sum_{i=1}^{I} v_i^1 x_{ip}^1 + \sum_{t=1}^{T} \mu_t^2 x_{tp}^2 = 1,$$
$$\sum_{k=1}^{K} v_k^3 x_{kp}^3 + \sum_{t=1}^{T} (v_t^2 - \mu_t^2) x_{tp}^2 + \sum_{d=1}^{D} w_d^1 z_{dp}^1 + \sum_{b=1}^{B} \varphi_b^2 z_{bp}^2 = 1,$$
$$\sum_{d=1}^{D} w_d^1 z_{dj}^1 + \sum_{b=1}^{B} w_b^2 z_{bj}^2 + \sum_{h=1}^{H} u_h^1 y_{hj}^1 - \sum_{i=1}^{I} v_i^1 x_{ij}^1 - \sum_{t=1}^{T} \mu_t^2 x_{tj}^2 \le 0, \quad \forall j = 1, ..., n,$$
$$\sum_{r=1}^{R} u_r^2 y_{rj}^2 - \sum_{k=1}^{K} v_k^3 x_{kj}^3 - \sum_{t=1}^{T} (v_t^2 - \mu_t^2) x_{tj}^2 - \sum_{d=1}^{D} w_d^1 z_{dj}^1 - \sum_{b=1}^{B} \varphi_b^2 z_{bj}^2 \le 0, \quad \forall j = 1, ..., n,$$
$$w_d^1, w_b^2, u_r^2, v_k^3, v_t^2, v_i^1, u_h^1 \ge 0, \quad \forall d = 1, .., D; b = 1, .., B; r = 1, .., s;$$
$$k = 1, ..K; t = 1, .., T; i = 1, .., I; h = 1, .., H$$
$$0 \le \varphi_b^2 \le w_b^2, \qquad \forall b = 1, ..., B,$$
$$0 \le \mu_t^2 \le v_t^2, \qquad \forall t = 1, ..., T,$$
$$\theta_{1p}^L \le \lambda \le \theta_{1p}^U \tag{20}$$

Now, consider the following transformations:

$$\tilde{w}_d^1 = \lambda w_d^1, \quad d = 1, ..., D,$$
$$\tilde{w}_b^2 = \lambda w_b^2, \quad b = 1, ..., B,$$
$$\tilde{u}_h^1 = \lambda u_h^1, \quad h = 1, ..., H,$$
$$\tilde{u}_r^2 = \lambda u_r^2, \quad r = 1, ..., R,$$
$$\tilde{v}_i^1 = \lambda v_i^1, \quad i = 1, ..., I,$$
$$\tilde{v}_t^2 = \lambda v_t^2, \quad t = 1, ..., T,$$
$$\tilde{v}_k^3 = \lambda v_k^3, \quad k = 1, ..., K,$$
$$\tilde{\varphi}_b^2 = \lambda \varphi_b^2, \quad b = 1, ..., B,$$
$$\tilde{\mu}_t^2 = \lambda \mu_t^2, \quad t = 1, ..., T$$

Table 1
Comparative overview of two-stage DEA models.

Method	Shared intermediate products	Shared inputs for two sub-DMUs	Direct input for Stage 2	Final output from Stage 1	Linear/ Non-linear	Parametric/ Non-parametric	Efficiency decomposition
Zhu (2000), Sexton and Levis (2003), Chen and Zhu (2004), Kao and Hwang (2008), Cook et al. (2010), Chen et al. (2010), Akther et al. (2013), Liu et al. (2015), Wang et al. (2016), Izadikhah and Farzipoor Saen (2016), Degl'Innocenti et al. (2017).	×	×	×	×	Linear	Non-parametric	✓
Chen et al. (2009), Wang and Chin (2010), Liu (2011a).	×	×	×	×	Linear	Non-parametric	✓
Zha and Liang (2010)	×	✓	×	×	Linear	Parametric	✓
Du et al. (2011)	×	×	×	×	Linear	Parametric	✓
Li et al. (2012), Guo et al. (2017).	×	×	✓	×	Linear	Parametric	✓
Yu and Shi (2014)	✓	×	✓	×	Linear	Non-parametric	✓
Toloo et al. (2015)	×	✓	×	×	Linear	Non-parametric	✓
Ma (2015)	✓	✓	✓	×	Linear	Non-parametric	×
Wu et al. (2016)	×	×	✓	✓	Linear	Parametric	✓
Huang, Chen and Lin (2017)	×	✓	×	×	Non-linear	-	×
Proposed method	✓	✓	✓	✓	Linear	Non-parametric	✓

By using the above transformations, Model (20) can be converted into the following linear programming problem:

$$
\begin{aligned}
&\theta_p = \max \quad \sum_{r=1}^{R} \tilde{u}_r^2 y_{rp}^2 \\
&s.t. \\
&\sum_{i=1}^{I} \tilde{v}_i^1 x_{ip}^1 + \sum_{t=1}^{T} \tilde{\mu}_t^2 x_{tp}^2 = \lambda, \\
&\sum_{k=1}^{K} \tilde{v}_k^3 x_{kp}^3 + \sum_{t=1}^{T} (\tilde{v}_t^2 - \tilde{\mu}_t^2) x_{tp}^2 + \sum_{d=1}^{D} \tilde{w}_d^1 z_{dp}^1 + \sum_{b=1}^{B} \tilde{\varphi}_b^2 z_{bp}^2 = \lambda, \\
&\sum_{d=1}^{D} \tilde{w}_d^1 z_{dj}^1 + \sum_{b=1}^{B} \tilde{w}_b^2 z_{bj}^2 + \sum_{h=1}^{H} \tilde{u}_h^1 y_{hj}^1 - \sum_{i=1}^{I} \tilde{v}_i^1 x_{ij}^1 - \sum_{t=1}^{T} \tilde{\mu}_t^2 x_{tj}^2 \le 0, \\
&\qquad\qquad \forall j = 1, ..., n, \\
&\sum_{r=1}^{R} \tilde{u}_r^2 y_{rj}^2 - \sum_{k=1}^{K} \tilde{v}_k^3 x_{kj}^3 - \sum_{t=1}^{T} (\tilde{v}_t^2 - \tilde{\mu}_t^2) x_{tj}^2 - \sum_{d=1}^{D} \tilde{w}_d^1 z_{dj}^1 - \sum_{b=1}^{B} \tilde{\varphi}_b^2 z_{bj}^2 \le 0, \\
&\qquad\qquad \forall j = 1, ..., n, \\
&\tilde{w}_d^1, \tilde{w}_b^2, \tilde{u}_r^2, \tilde{v}_k^3, \tilde{v}_t^2, \tilde{v}_i^1, \tilde{u}_h^1 \ge 0, \quad \forall d = 1, .., D; b = 1, .., B; r = 1, .., s; \\
&\qquad\qquad k = 1, ..K; t = 1, .., T; i = 1, .., I; h = 1, .., H \\
&0 \le \tilde{\varphi}_b^2 \le \tilde{w}_b^2, \qquad \forall b = 1, ..., B, \\
&0 \le \tilde{\mu}_t^2 \le \tilde{v}_t^2, \qquad \forall t = 1, ..., T, \\
&\theta_{1p}^L \le \lambda \le \theta_{1p}^U.
\end{aligned}
\tag{21}
$$

By solving Model (21), we obtain the optimal value θ_p^* of the overall efficiency of the cooperative model. Model (21) also shows that the optimal value of λ coincides with the efficiency score of Stage 1 of DMU_p, i.e. $\theta_{1p}^* = \lambda$. Finally, the efficiency score of Stage 2 of DMU_p can be calculated as $\theta_{2p}^* = \theta_p^*/\theta_{1p}^*$.

On the other hand, after obtaining the values of $\tilde{\mu}_t^2$ and $\tilde{v}_t^2$ from Model (21), we can calculate the value of the allocative factor α_t ($t = 1, ..., T$) as follows: $\alpha_t = \tilde{\mu}_t^2/\tilde{v}_t^2$. Similarly, the values of $\tilde{\varphi}_b^2$ and $\tilde{w}_b^2$ ($b = 1, ..., B$) obtained by solving Model (21) can be used to calculate the value of the allocative factor β_b as follows: $\beta_b = \tilde{\varphi}_b^2/\tilde{w}_b^2$.

5. Proposed method versus existing methods

There are many novel performance measurement studies in the two-stage DEA literature. Table 1 presents a comparative overview of the most widely used two-stage DEA models outlining the advantages provided by the method introduced in this study when compared to them.

Our method is more general than the other existing models. It allows for many more possibilities when assigning direct inputs and final outputs and, consequently, for a wider range of applications. In order to support this statement in a more concrete manner, we performed a numerical comparison between our method and the three reference methods used to define it, that is, the methods proposed by Kao and Hwang (2008), Zha and Liang (2010) and Yu and Shi (2014).

To obtain a meaningful comparison, we used two sets of data. The first dataset is taken from Kao and Hwang (2008) and is shown in Table 2. This set includes the input and output data for three DMUs denoted by A, B and C. Each DMU is characterized by one initial input, one intermediate product, and one output.

The results from running the three aforementioned models along with those produced by our method are illustrated in Table 3. These results confirm the soundness of our model and its capacity of reliably solving the two-stage DEA problem.

The second dataset is taken from Yu and Shi (2014). The results reported in Table 4 show the closeness and consistency of the solutions obtained by implementing the proposed method with respect to those obtained by applying Yu and Shi's method.

Table 2
Dataset taken from Kao and Hwang (2008).

DMU	Input	Intermediate product	Output
A	2	1.5	1.5
B	4	4	5
C	5	5	6

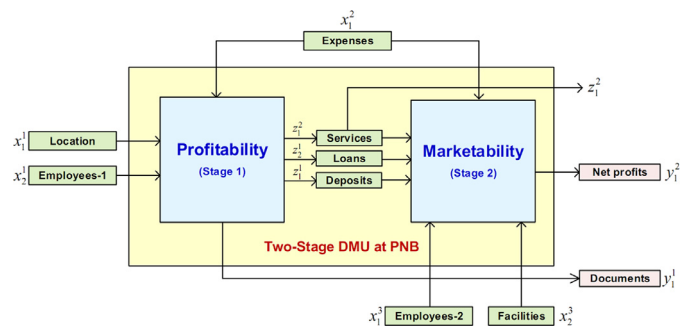

Fig. 5. Proposed two-stage structure of a branch of the PNB system.

Finally, it must be underlined that, compared to other existing methods, our approach is capable of analyzing DMUs and the efficiency of their single sub-processes in a much larger number of real-life situations thanks to the much more general conditions assumed for the DMUs in terms of inputs, intermediate products and outputs. In particular, the fact that it allows for shared initial inputs and final outputs to be produced in Stage 1 makes it more flexible and operatively more complete than Yu and Shi's method. Case studies such as the one considered in the next section can neither be solved by using the model of Kao and Hwang (2008), Zha and Liang (2010) or Yu and Shi (2014), nor by running many of the recently developed models. The structure assumed for the DMUs needs to allow for more complex and interlinking possibilities as it is the case in our setting.

6. Case study

Defining performance evaluation procedures for commercial banks constitutes one of the most important and promising application areas of DEA. In this section, we apply the proposed method to evaluate 15 branches of the Philadelphia National Bank (PNB)[1] in the Philadelphia metropolitan area.

Seiford and Zhu (1999) investigated the efficiencies of a set of listed U.S. commercial banks by interpreting them as two-stage processes: Stage 1 was evaluated on the profitability performance while Stage 2 focused on the marketability performance. Following Seiford and Zhu (1999), we model the PNB as a two-stage production system with freely distributed inputs and shared intermediate outputs. Each branch consists of two stages, with Stage 1 representing *profitability* and Stage 2 representing *marketability*. Fig. 5 provides a graphic summary of the two-stage structure proposed to analyze the PNB system.

It should be noted that Seiford and Zhu (1999) – and Liu (2011b) – analyzed the capacity of a bank to generate revenue and profit as outputs following from both the profitability and marketability stages. In particular, the revenue and profit obtained from the profitability stage follow from the operational capacity of the bank or financial company (i.e. the efficient use of labor, assets,

[1] The names of the bank branches have been changed to protect the anonymity of the bank.

Table 3
Numerical comparison among efficiency evaluation methods.

DMU	Stage 1				Stage 2				Overall			
	KH	ZL	YS	Proposed method	KH	ZL	YS	Proposed method	KH	ZL	YS	Proposed method
A	0.75	0.75	0.80	0.75	0.80	0.80	0.80	0.80	0.60	0.48	0.60	0.60
B	1	1	1	1	1	1	1	1	1	1	1	1
C	1	1	0.96	1	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96

KH: Kao and Hwang (2008), **ZL:** Zha and Liang (2010), and **YS:** Yu and Shi (2014)

Table 4
Numerical comparison between the proposed method and Yu and Shi (2014).

DMU	Proposed method			Yu and Shi (2014)		
	Overall	TETA_1	TETA_2	Overall	TETA_1	TETA_2
1	0.894	1	0.894	0.896	0.957	0.936
2	0.612	1	0.612	0.612	1	0.612
3	0.800	0.800	1	0.800	0.800	1
4	0.628	1	0.628	0.628	1	0.628
5	0.408	0.676	0.604	0.408	0.676	0.604
6	1	1	1	1	1	1
7	0.833	1	0.833	0.833	1	0.833
8	0.513	0.770	0.666	0.770	0.772	1
9	0.500	0.500	1	0.500	0.500	1
10	0.596	1	0.596	0.596	1	0.596

and capital resources) while those derived from the marketability stage follow from its performance in the stock market. Thus, profits and revenues are obtained in both stages and serve also as intermediate factors between them. It should be noted that the factors considered by these authors to measure the overall performance of a bank across both stages are taken from its financial report.

In the current paper, we must modify the profitability and marketability framework introduced by Seiford and Zhu (1999) and adapt it to the operational process of the branches of a commercial bank. The working of bank branches, based on a direct and personal interaction with their customers, determines their capacity to generate profits. As an example of the fundamental role of this process consider, for example, the work of Menon and O'Connor (2007), who highlight the importance that the customer relationship management strategy of banks has to trigger the affective commitment of customers through their interactions with bank employees. Indeed, this type of commitment determines the capacity of the branch to generate profits. The two-stage case study analyzed in the current paper operates at this level of interaction.

We adapt the two-stage division of a commercial bank to account for the standard working procedure of its branches. In this regard, the profitability stage described in Fig. 5 reflects the direct interaction with customers that takes place when providing different types of banking services, ranging from the collection of deposits to the concession of loans. The services provided can be either final or part of more complex financial operations that must be developed further. This initial stage is required for the clients to access the branch, interact at the front desk, deposit funds, apply for loans, or perform any other financial operation. Different types of documents – together with several final services – will be produced as a consequence of these personal interactions. This operational stage determines the capacity of the bank to obtain initial revenues that will be used in the second stage to generate net profits.

In this regard, the marketability stage describes the different procedural operations undertaken within the branch. The provision and development of complex financial intermediation services, together with the management of deposits and loans, constitute the procedural or bureaucratic stage of the branch. That is, the financial interactions and bookkeeping procedures guaranteeing the correct operation of the facilities when performing further retrieval and deposit operations determine the capacity of the branch to generate profits –after accounting for the different operational costs incurred through both stages.

It follows from the above description that the choice of inputs and outputs is an important task since it determines the degree of coherency and reliability of the efficiency scores obtained through the selected measuring method. More precisely, the choice of inputs and outputs must reflect the mission of the bank under analysis.

There is a large number of studies dealing with the efficiency measurement problem in the banking industry. The variables to be used as inputs and outputs in the case study were suggested by the senior management at PNB upon the recommendation of a panel of experts. The experts were asked to determine the most appropriate and significant evaluation criteria on the basis of the available data and their relevance to the bank operations. These criteria are shown in Table 5 together with the relevant literature that was used to identify them. Table 6 describes the allocation of data in terms of the inputs, intermediate outputs and outputs used for evaluating the PNB branches.

The inputs of Stage 1 are: Location (x_1^1), which is expressed as a score between 0 and 5 assigned to accessibility, security, and parking availability by the customers of each branch (low scores are given to good locations and high scores are given to poor locations); Employees-1 (x_2^1) corresponding to the number of employees in Stage 1; and Expenses (x_1^2) representing the total annual operating expenses of the branch. Expenses are also considered in Stage 2. The inputs of Stage 2 are: Employees-2 (x_1^3) accounting for the number of employees in Stage 2; and Facilities (x_2^3), which are represented as scores between 0 and 60 assigned to the number of ATM machines, waiting time, change counting machines, amenities in the waiting area, and access to Wi-Fi by the customers of each branch. The intermediate products are: Deposits (z_1^1) representing short-term and long-term deposits; Loans (z_2^1) standing for the customer and business loans; and Services (z_1^2), which represent the diversity of the bank services. We account for the total number of service operations performed within a branch by considering the different ways in which a bank helps customers, which include transfer payments, account operations, the payment of standing orders and exchanging foreign currency. Some parts of services are also considered as outputs in Stage 2. The outputs of Stage 2 are: Documents (y_1^1), which represent the amount of cash, transfer, and promissory notes; and Net Profits (y_1^2) indicating the overall profits of the branch. Note that each stage has a different number of Employees. The employees in Stage 1 work in the profitability group focusing on generating profits while the employees in Stage 2 work in the marketability group focusing on new markets and the production of revenues and sales.

The dataset used in the case study is provided in Table 7. The data refer to the period 2014–2015. Table 8 provides the characteristics and main descriptive statistics on the inputs and outputs of the 15 PNB branches. The first column lists the names of the variables while columns 2 through 5 reports various statistics includ-

Table 5
Performance measurement criteria used in the case study.

Employee	Huang, Chen et al. (2017), Wanke et al. (2016), Huang, Chen et al. (2017), Chen, Chiu, Huang, and Tu (2013), Akther et al. (2013), Wang et al. (2014), Ray (2016), Zhou et al. (2017), Ding et al. (2017), Degl'Innocenti et al. (2017), Chortareas et al. (2016), Wanke and Barros (2014), Wanke et al. (2017), Fukuyama & Matousek (2017), Kamarudin et al. (2017) and Widiarto et al. (2017).
Deposits	Chen, Liang, and Zhu (2009), Hosseinzadeh Lotfi et al. (2012), Sun et al. (2013), Huang, Chen et al. (2017), Stewart et al. (2016), Wanke et al. (2016), Tavana and Khalili-Damghani (2014), Chen et al. (2013), Akther et al. (2013), Wang et al. (2014), Ray (2016), Zhou et al. (2017), Degl'Innocenti et al. (2017), Silva et al. (2017), Chortareas et al. (2016), Herrera-Restrepo et al. (2016), Wanke et al. (2017), Fukuyama & Matousek (2017), Kamarudin et al. (2017) and San-Jose, Retolaza, and Torres Pruñonosa (2014).
Loans	Huang, Chen et al. (2017), Stewart et al. (2016), Huang, Chen et al. (2017), Chen et al. (2013), Akther et al. (2013), Zhou et al. (2017), Degl'Innocenti et al. (2017), Silva et al. (2017), Chortareas et al. (2016), Wanke et al. (2017), Fukuyama & Matousek (2017), Kamarudin et al. (2017), Widiarto et al. (2017).
Expenses	Du et al. (2014), Sun et al. (2013), Liu (2014a,b), Chen et al. (2009), Sun et al. (2013), Tavana and Khalili-Damghani (2014), Avkiran (2015), Zhou et al. (2017), Silva et al. (2017), Wanke and Barros (2014), Afsharian et al. (2016), Widiarto et al. (2017).
Services	Tavana and Khalili-Damghani (2014), Degl'Innocenti et al. (2017), Chortareas et al. (2016), Herrera-Restrepo et al. (2016), San-Jose et al. (2014).
Facilities	Diallo (2017), Degl'Innocenti et al. (2017), Chortareas et al. (2016), Herrera-Restrepo et al. (2016), San-Jose et al. (2014).
Location	Diallo (2017), Herrera-Restrepo et al. (2016).
Documents	Diallo (2017), Herrera-Restrepo et al. (2016).
Net profit	Sahin et al. (2016), Wanke et al. (2017), Ebrahimnejad et al. (2014), Tavana and Khalili-Damghani (2014), Wang et al. (2014), Avkiran (2015).

Table 6
Inputs, intermediates, and outputs of the two-stage branches of the case study.

Input	Stage 1 input	x_1^1: Location
	Stage 1 input	x_2^1: Employees-1
	Stage 1 input shared with Stage 2	x_1^2: Expenses
	Stage 2 input	x_1^3: Employees-2
	Stage 2 input	x_2^3: Facilities
Intermediate	Stage 1 output – Stage 2 input	z_1^1: Deposits
	Stage 1 output – Stage 2 input	z_2^1: Loans
	Stage 1 output – Stage 2 input and final output	z_1^2: Services
Output	Stage 1 output	y_1^1: Documents
	Stage 2 output	y_1^2: Net profits

ing the mean, standard deviation, minimum value, and maximum value.

The high standard deviation for one of the input variables, i.e. 'Expenses', in Table 8 suggests the presence of big differences across the existing banks. These differences lead to high standard deviation values for 'Deposits' and 'Loans'. The reason behind this fact is that some banks which consume large 'Expenses' are more specialized in lending activities and tend to have more loans compared to other banks (Degl'Innocenti et al., 2017). Therefore, it is plausible to find a higher standard deviation for intermediate output variables such as 'Deposits' and 'Loans'. A similar reasoning applies to the variables of the second stage.

Table 9 reports the results obtained by applying Models (10) to (17). The second and third columns show the upper and lower bounds of non-cooperative efficiency for Stage 1; the fourth and fifth columns show the upper and lower bounds of non-cooperative efficiency for Stage 2. Note, in particular, that the Manayunk branch is efficient in both its stages and in both non-cooperative cases since all the four efficiency bounds take the value of 1.

Table 10 reports the cooperative efficiencies and the relative efficiencies of the two stages. The fourth column provides the global optimal values obtained by solving Model (21). The sixth and seventh columns represent the optimal proportion of the shared inputs and intermediate outputs that must be allocated to Stage 1 and Stage 2, respectively.

From Table 10, we can conclude that 47% of the PNB branches are efficient in Stage 1, 20% of them are efficient in Stage 2, and, finally, 7% of them become overall efficient. Fig. 6 illustrates the overall efficiency values obtained for all the DMUs.

The proposed two-stage DEA models were coded using GAMS 23.6 software. As Table 10 shows, seven bank branches result to be efficient in the first stage, which means that there is no efficiency loss during the initial operational stage, three bank branches are credited to be efficient in the second stage, but only one bank branch, the Manayunk branch, works efficiently in both stages. Therefore, only one of the 15 bank branches under analysis is overall efficient.

The efficiency values θ_1^* obtained in Stage 1 lie in the interval [0.471622, 1] and are in general close to each other, delivering a mean value of 0.869902 and a standard deviation of 0.157961. Therefore, the performance of the bank branches in the profitability stage is sufficiently high so as to deliver a global performance evaluation that could be considered adequate enough. On the other hand, the performance of the branches in the marketability stage is considerably lower. In particular, the efficiency scores θ_2^* obtained in Stage 2 are completely diffused with a mean value of 0.347636 and a standard deviation of 0.41519.

The sixth column of Table 10 represents the optimal proportion of the freely distributed inputs "Expenses" that must be allocated in Stage 1. For example, the branch expenses of Aramingo are to be entirely allocated in its first stage. Also, 35.4 % of the branch expenses of Belmont must be allocated in its first stage and the remaining 64.6 % in the second stage.

The seventh column of Table 10 represents the optimal proportion of the intermediate outputs "Services" that must be allocated in Stage 2. For example, 0.7 % of the branch services of Richmond are to be allocated in its second stage and the remaining 99.3 % considered as final outputs of the Richmond branch.

Finally, note that the Roxborough and Manayunk branches resulted to be the worst and the best branches, respectively. The

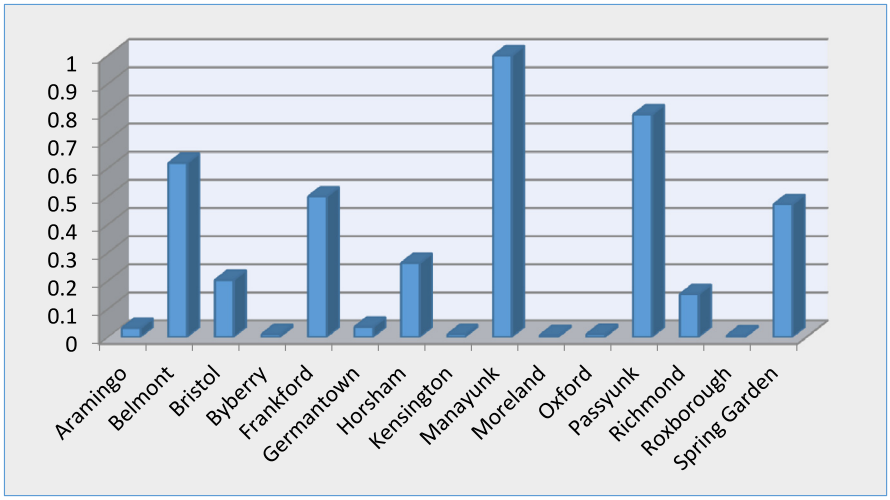

Fig. 6. Comparison among the overall efficiency values obtained in the case study.

ranking obtained for all the DMUs is shown in the fifth column of Table 10.

6.1. Comparing our efficiency results with those provided by a black-box approach

For comparison purposes, we also estimated the performance of the 15 bank branches considered in the case study using a black box technology that does not account for the intermediate products; see Table 11.

To obtain a valid trend for the evaluation of the PNB system when the DMUs are regarded as black-boxes, we have considered three cases:

(a) $\beta = 0$, i.e. all the branch services (z_1^2) are treated as final output;
(b) $\beta = 0.5$, i.e. half of the branch services are treated as final output;
(c) $\beta = 1$, i.e. no amount of branch services is treated as final output.

As shown in Table 11, at least 60 % of the DMUs are efficient in all the three cases. Thus, the results for the overall performances in the black-box approach (see the Black-Box column of Table 11) are very far from those obtained by the proposed model based on an intermediate-like approach (see the fourth column of Table 10 and the Proposed Two-Stage column of Table 11).

The differences between the overall performance results clearly indicate the fact that a black-box efficiency evaluation procedure may not properly represent the aggregate performances of the DMUs composing the system.

Remark. Note that there is one DMU, i.e. “Roxborough”, whose stages are both inefficient in the proposed two-stage model despite being always overall efficient in the black box model. This fact is considered by Kao and Hwang (2008) as a signal of inconsistency and inadequacy when evaluating the overall efficiency using the black box model. On the contrary, Xiang and Li (2017) do not see this type of conflicts as a critique to the black box approach but as a matter of correctly interpreting the concepts of externality and coordination with respect to the system internal organization. In other words, the presence of a DMU that is inefficient in both its stages but overall efficient when no sub-process is considered would represent only an apparent conflict that can be explained by a relatively better coordination within its structure. ■

Remark. Regarding the dataset collection, one may fairly object that banking efficiency studies usually use a larger dataset than the one implemented in this case study, namely, a dynamic dataset accounting for multiple time periods. However, this would require the definition of a dynamical framework that, while being an interesting and recommendable extension of the proposed model, does not constitute the main objective of the present paper. As underlined earlier (see Sections 1 and 5), the focus of this study is on the definition of a two-stage structure capable of overcoming several drawbacks of the existing models by allowing for stronger and more general assumptions on the DMUs in terms of inputs, intermediate products and outputs. Thus, our case study aims at showing the goodness of our assumptions and how well-behaved is the developed mathematical model when it comes to real-life situations that cannot be analyzed using any of the exiting methods. ■

6. Conclusion and future research directions

Conventional DEA models interpret DMUs as black-boxes that consume a set of inputs to produce a set of outputs without taking into consideration the intermediate performance measures that characterize a DMU. That is, traditional DEA studies view a system as a sole block and ignore the performance of the internal processes in calculating the relative efficiency within a production system. As a result, intermediate measures are lost in the process of transforming the inputs into outputs and it becomes difficult, if not impossible, to provide individual DMU managers with specific information on the components of a DMU responsible for the overall inefficiency.

In this sense, independently from the level of generality guaranteed by the assumptions of the model under analysis, many studies have pointed out the existence of contradictory results when evaluating DMUs via a two-stage and a black-box approach. More often than not, the case studies proposed to validate new approaches have shown the concrete possibility for a DMU to be overall efficient as a black-box even though it is inefficient in both its sub-processes when regarded as a two-stage structure (Kao & Hwang, 2008; Xiang & Li, 2017). For example, this happens to be true for one of the DMUs analyzed in our case study (i.e., the bank branch referred to as “Roxborough”).

Ignoring the internal structure of a process by treating it as a black box is less than optimal and often unreliable since it prevents one from identifying the real sources of inefficiency and may, consequently, lead to inaccurate efficiency evaluations (Rho & An, 2007; Ebrahimnejad et al., 2014).

Table 7
Dataset used in the case study.

DMU	x_1^1 Location	x_2^1 Employees-1	x_1^2 Expenses	x_1^3 Employees-2	x_2^3 Facilities	z_1^1 Deposits	z_2^1 Loans	z_1^2 Services	y_1^1 Documents	y_1^2 Net profits
Aramingo	4	33	6,053,950	31	45	88,805,870	9,930,070	1586	202,402	886,700
Belmont	5	42	2,568,000	41	48	80,331,040	10,558,070	2587	130,134	12,504,200
Bristol	3	31	4,937,460	20	39	54,096,730	26,908,410	1928	157,240	3883,930
Byberry	5	38	13,770,420	51	58	159,553,230	23,552,210	4334	316,914	369,000
Frankford	5	34	16,406,000	52	47	176,834,230	48,220,890	5051	334,180	13,600,000
Germantown	3	44	5,602,820	19	54	78,190,480	22,086,900	3417	162,561	482,030
Horsham	4	46	2,961,270	28	56	64,369,880	27,478,490	2533	263,767	3,388,480
Kensington	4	33	2,566,360	41	37	59,637,190	18,286,690	3141	236,540	118,910
Manayunk	4	43	2,231,200	63	57	116,929,340	13,285,0050	4558	246,474	16,330,000
Moreland	3	34	6,102,760	30	41	60,218,920	16,808,870	3161	240,654	122,670
Oxford	4	41	11,888,260	47	53	124,104,230	32,805,910	3196	221,537	311,640
Passyunk	3	31	5,122,280	29	34	66,308,610	17,785,790	2548	162,928	19,708,980
Richmond	4	34	3,943,260	34	43	69,911,600	43,754,650	2518	286,015	2,500,000
Roxborough	4	31	4,451,080	22	44	46,755,500	19,889,450	2266	212,533	102,780
Spring Garden	4	40	4,075,000	16	51	28,471,320	31,604,330	1286	131,854	11,610,000

Table 8
Descriptive statistics of the two-stage branches of the case study.

Variable	Mean	Std. Dev.	Min	Max
x_1^1: Location	3.933333	0.703732	3	5
x_2^1: Employees-1	37	5.237229	31	46
x_1^2: Expenses	6,178,675	4,328,220	2,231,200	16,406,000
x_1^3: Employees-2	34.93333	13.84334	16	63
x_2^3: Facilities	47.13333	7.567474	34	58
z_1^1: Deposits	84,967,878	41,829,689	28,471,320	1.77E+08
z_2^1: Loans	32,168,052	29,884,038	9,930,070	1.33E+08
z_1^2: Services	2940.667	1071.356	1286	5051
y_1^1: Documents	220,382.2	63,701.57	130,134	334,180
y_1^2: Net profits	5,727,955	6,934,316	102,780	19,708,980

Table 9
Upper and lower bounds of non-cooperative efficiency obtained in the case study.

DMU	θ_1^L	θ_1^U	θ_2^L	θ_2^U
Aramingo	0.714	0.826	0.042	0.080
Belmont	0.576	0.690	0.896	1
Bristol	0.690	0.729	0.278	0.278
Byberry	0.917	0.953	0.011	0.014
Frankford	1	1	0.499	0.499
Germantown	1	1	0.035	0.036
Horsham	1	1	0.263	0.263
Kensington	1	1	0.010	0.010
Manayunk	1	1	1	1
Moreland	1	1	0.007	0.007
Oxford	0.757	0.895	0.010	0.012
Passyunk	0.791	0.791	1	1
Richmond	1	1	0.152	0.152
Roxborough	0.834	0.834	0.007	0.007
Spring Garden	0.472	0.472	1	1

Note: This table reports the upper and lower bounds of non-cooperative efficiency obtained in the case study for the 15 branches of the Philadelphia National Bank in the Philadelphia metropolitan area. θ_1^L and $\theta_1^U$are the lower bound and upper bounds for Stage 1, respectively; θ_2^L and θ_2^U are the lower and upper bounds for Stage 2, respectively.

Table 10
Efficiency results delivered by the proposed model in the case study.

DMU	θ_1^*	θ_2^*	θ^*	**Ranking**	α_1	β_1
Aramingo	0.773	0.042	0.032	10	1	1
Belmont	0.690	0.896	0.618	3	0.354	1
Bristol	0.729	0.278	0.203	7	0.354	1
Byberry	0.917	0.011	0.010	13	1	1
Frankford	1	0.499	0.499	4	1	1
Germantown	1	0.036	0.036	9	0.432	0
Horsham	1	0.263	0.263	6	0.275	0
Kensington	1	0.010	0.010	12	0.113	0
Manayunk	1	1	1	1	0.113	0
Moreland	1	0.007	0.007	14	0	0
Oxford	0.844	0.012	0.010	11	0	0
Passyunk	0.791	1	0.791	2	0	0
Richmond	1	0.152	0.152	8	0.233	0.007
Roxborough	0.834	0.007	0.006	15	0.233	0.007
Spring Garden	0.472	1	0.472	5	0.233	0
Mean	0.870	0.348	0.274			
Std. Dev.	0.158	0.415	0.326			
Min	0.472	0.007	0.006			
Max	1	1	1			

Note: This table reports the efficiency results obtained by implementing the proposed method for the 15 branches of the Philadelphia National Bank in the Philadelphia metropolitan area. θ_1^* is the efficiency of Stage 1; θ_2^* is the efficiency of Stage 2; θ^* is the overall efficiency; α_1 is the portion of x_{tj}^2 consumed in Stage 1; β_1 is the portion of z_{bj}^2 consumed in Stage 2.

Table 11
Proposed method efficiency scores vs black-box approach efficiency scores.

DMU	Proposed two-stage overall efficiency	Black-Box $\beta=0$	$\beta=0.5$	$\beta=1$
Aramingo	0.032	0.758	0.758	0.758
Belmont	0.618	0.896	0.896	0.896
Bristol	0.203	0.922	0.922	0.893
Byberry	0.010	0.917	0.917	0.917
Frankford	0.499	1	1	1
Germantown	0.036	1	1	0.899
Horsham	0.263	1	1	1
Kensington	0.010	1	1	1
Manayunk	1	1	1	1
Moreland	0.007	1	1	1
Oxford	0.010	0.752	0.752	0.719
Passyunk	0.791	1	1	1
Richmond	0.152	1	1	1
Roxborough	0.006	1	1	1
Spring Garden	0.472	1	1	1

Note: This table reports the overall efficiency scores produced by the black-box approach for the 15 bank branches of the case study and compares them with ours. This comparison is carried out considering three values of β, i.e. the portion of 'branch services' that is consumed in Stage 2.

In other words, a black-box approach to efficiency scores may not properly represent the aggregate performances inherent to the processes composing a system. Thus, the objective of efficiency measurement becomes to detect the weak points of a system so that appropriate efforts can be devoted to improve its performance. Regarding, in particular, inefficient DMUs, an issue of great concern is to establish what factors cause the inefficiency.

To answer this kind of questions, it is necessary to break down the overall efficiency into components so as to identify the sources of inefficiency. One way to proceed with such a decomposition is focusing on the internal structure of the DMUs through DEA models.

In this paper, we have considered a novel two-stage DEA model where each DMU is composed by two sub-DMUs in series, the intermediate products of the sub-DMU in Stage 1 are partly consumed by the sub-DMU in Stage 2, and the initial inputs of the DMU can be freely allocated to either one of the two sub-DMUs. Also, there are additional inputs directly consumed in Stage 2 while the sub-DMU in Stage 1 is allowed to produce final outputs.

Building on the concept of Stackelberg leader–follower game and the work of Yu and Shi (2014), we formulated four fractional programming problems that allow to determine the upper and lower bounds of the efficiencies of the two sub-DMUs in the non-cooperative setting. Afterwards, we have considered the cooperative setting and defined a fractional programming model to calculate the overall efficiency of the DMUs. All the models have been linearized by applying suitable Charnes–Cooper transformations (Charnes & Cooper, 1962).

The proposed two-stage DEA model allows for important applications to several management areas. We have considered an application to banking. In a banking environment, there is always a need for tools allowing one to uncover the inefficiencies that can affect the different components of an operation.

Banking operations and production processes are often formulated as two-stage processes (Zha & Liang, 2010; Matthews, 2013; Akther et al., 2013). While in the traditional one-stage DEA methods deposits are treated as an input or as an output, in two-stage DEA models they are considered intermediate variables and generally used as outputs in Stage 1 and inputs in Stage 2. Consequently, the dual role of the deposits is kept intact (Wanke & Barros, 2014; Liu, 2014a,b; Ebrahimnejad et al., 2014; Degl'Innocenti et al., 2017).

Seiford and Zhu (1999) proposed the first two-stage DEA model to evaluate the profitability and marketability of a group of commercial banks in the United States. They evaluated the single stage efficiencies independently by implementing a standard DEA method. The inconsistency between the results obtained for the entire process and those relative to its components revealed the incapacity of their model to explain the interrelationships existing within the two-stage structures considered (Wang et al., 2014; Avkiran, 2015). Nonetheless, their pilot study has laid a foundation for multi-stage production modeling to be applied to banks and other similar industries. Among the most recent studies interpreting banks as multi-stage production units, we find Ohsato and Takahashi (2015), Fukuyama and Matousek (2017), Huang, Lin, et al. (2017), Wu et al. (2016), Zhou et al. (2017).

In the case study section, the proposed model has been applied to evaluate the efficiency of 15 commercial bank branches in the U.S. According to the results obtained, 14 out of the 15 bank branches analyzed turned out to be overall inefficient. Only one bank branch was recognized as efficient during the 2014–2015 period examined. This implies that mostly all the branches need to actively spur innovation and future growth by identifying and developing productive investment and appropriate management activities.

In the first stage, the sub-process of profitability measurement, seven branches were identified as efficient. In the second stage, the sub-process of marketability measurement, three branches were recognized to perform efficiently. From a statistical viewpoint, the efficiency of the first stage is higher than that of the second stage. This indicates that the low efficiency scores obtained for the two-stage processes were mainly due to the low efficiency scores of the corresponding second stages, that is, the marketability efficiency values. In particular, this confirmed the finding of Luo (2003) who sees in the marketability stage, rather than in the profitability one, the real cradle of bank inefficiency problems.

Finally, for comparison purposes, the overall performances of the 15 bank branches were also estimated using a black-box technology that does not account for the intermediate products. The results delivered by the black-box approach turned out to be considerably far from those obtained via the intermediate-like approach characterizing the proposed model. This allows us to state the inadequacy of black-box efficiency evaluation procedures to properly represent the aggregate performances of the DMUs composing a system.

We conclude with a few possible research directions towards which to extend the results of this study. The proposed model assumed constant returns to scale: it would be interesting to work out an extension of the model that applies to VRS technologies. The proposed model built on the key idea that the overall efficiency of a DMU can be decomposed in the product of the efficiencies of its two sub-DMUs (Kao & Hwang, 2008): future research is expected to consider an additive decomposition of the overall efficiency of a DMU. We also propose applying the model developed to measure efficiency in other research fields such as regional R&D processing, evaluating non-life insurance companies, and so on. In addition, our approach could be extended to model general series systems with multi-stage DMUs or to a dynamic setting.

Acknowledgement

The authors would like to thank the anonymous reviewers and the editor for their insightful comments and suggestions.

References

Akther, S., Fukuyama, H., & Weber, W. L. (2013). Estimating two-stage network Slacks-based inefficiency: An application to Bangladesh banking. *Omega, 41*(1), 88–96. http://dx.doi.org/10.1016/j.omega.2011.02.009.

Amin, G. R., & Toloo, M. (2004). A polynomial-time algorithm for determining the non-Archimedean epsilon in DEA models. *Computers & Operations Research, 31*(5), 803–805.

Avkiran, N. K. (2015). An illustration of dynamic network DEA in commercial banking including robustness tests. *Omega, 55*, 141–150. http://dx.doi.org/10.1016/j.omega.2014.07.002.

Castelli, L., Pesenti, R., & Ukovich, W. (2010). A classification of DEA models when the internal structure of the decision making units is considered. *Annals of Operations Research, 173*(1), 207–235. doi:10.1007/s10479-008-0414-2.

Charnes, A., & Cooper, W. W. (1962). Programming with linear fractional functionals. *Naval Research Logistics Quarterly, 9*, 181–186.

Charnes, A., Cooper, W. W., Golany, B., Halek, R., Klopp, G., Schmitz, E., et al. (1986). Two phase data envelopment analysis approach to policy evaluation and management of army recruiting activities: Tradeoffs between joint services and army advertising *research report CCS no.* 532. *Texas: Center for cybernetic studies*. The University of Texas at Austin.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research, 2*(6), 429–444. http://dx.doi.org/10.1016/0377-2217(78)90138-8.

Chen, Y.-C., Chiu, Y.-H., Huang, C.-W., & Tu, C. H. (2013). The analysis of bank business performance and market risk—Applying Fuzzy DEA. *Economic Modelling, 32*, 225–232. http://dx.doi.org/10.1016/j.econmod.2013.02.008.

Chen, Y., Cook, W. D., Li, N., & Zhu, J. (2009). Additive efficiency decomposition in two-stage DEA. *European Journal of Operational Research, 196*(3), 1170–1176. http://dx.doi.org/10.1016/j.ejor.2008.05.011.

Chen, Y., Cook, W. D., & Zhu, J. (2010). Deriving the DEA frontier for two-stage processes. *European Journal of Operational Research, 202*(1), 138–142. http://dx.doi.org/10.1016/j.ejor.2009.05.012.

Chen, Y., Liang, L., Yang, F., & Zhu, J. (2006). Evaluation of information technology investment: A data envelopment analysis approach. *Computers & Operations Research, 33*(5), 1368–1379. http://dx.doi.org/10.1016/j.cor.2004.09.021.

Chen, Y., Liang, L., & Zhu, J. (2009). Equivalence in two-stage DEA approaches. *European Journal of Operational Research, 193*(2), 600–604. http://dx.doi.org/10.1016/j.ejor.2007.11.040.

Chen, Y., & Zhu, J. (2004). Measuring information technology's indirect impact on firm performance. *Information Technology and Management, 5*(1-2), 9–22. doi:10.1023/B:ITEM.0000008075.43543.97.

Chortareas, G., Kapetanios, G., & Ventouri, A. (2016). Credit market freedom and cost efficiency in US state banking. *Journal of Empirical Finance, 37*, 173–185. http://dx.doi.org/10.1016/j.jempfin.2016.03.002.

Cook, W. D., Liang, L., & Zhu, J. (2010). Measuring performance of two-stage network structures by DEA: A review and future perspective. *Omega, 38*(6), 423–430. http://dx.doi.org/10.1016/j.omega.2009.12.001.

Degl'Innocenti, M., Kourtzidis, S. A., Sevic, Z., & Tzeremes, N. G. (2017). Bank productivity growth and convergence in the European Union during the financial crisis. *Journal of Banking & Finance, 75*, 184–199. http://dx.doi.org/10.1016/j.jbankfin.2016.11.016.

Despotis, D., Koronakos, G., & Sotiros, D. (2012). Additive decomposition in two-stage DEA: An alternative approach. *MPRA: Munich Personal RePEc Archive, 41724*, 1–23.

Diallo, B. (2017). Bank efficiency and industry growth during financial crises. *Economic Modelling*. http://dx.doi.org/10.1016/j.econmod.2017.03.011.

Ding, J., Dong, W., Liang, L., & Zhu, J. (2017). Goal congruence analysis in multi-division organizations with shared resources based on data envelopment analysis. *European Journal of Operational Research, 263*(3), 961–973. http://dx.doi.org/10.1016/j.ejor.2017.06.040.

Du, J., Chen, Y., Cook, W., Liang, L., & Zhu, J. (2014). Evaluating two-stage network structures: Bargaining game approach. In W. D. Cook, & J. Zhu (Eds.). In *Data envelopment analysis: 208* (pp. 165–187). US: Springer.

Du, J., Liang, L., Chen, Y., Cook, W. D., & Zhu, J. (2011). A bargaining game model for measuring performance of two-stage network structures. *European Journal of Operational Research, 210*(2), 390–397. http://dx.doi.org/10.1016/j.ejor.2010.08.025.

Ebrahimnejad, A., Tavana, M., Lotfi, F. H., Shahverdi, R., & Yousefpour, M. (2014). A three-stage data envelopment analysis model with application to banking industry. *Measurement, 49*(0), 308–319. http://dx.doi.org/10.1016/j.measurement.2013.11.043.

Färe, R., & Grosskopf, S. (2000). Network DEA. *Socio-Economic Planning Sciences, 34*, 35–49.

Färe, R., & Grosskopf, S. (1996). Productivity and intermediate products: A frontier approach. *Economics Letters, 50*(1), 65–70. http://dx.doi.org/10.1016/0165-1765(95)00729-6.

Färe, R., & Whittaker, G. (1995). An intermediate input model of dairy production using complex survey data. *J Agric Econ, 46*(2), 201–213.

Fukuyama, H., & Matousek, R. (2017). Modelling bank performance: A network DEA approach. *European Journal of Operational Research, 259*(2), 721–732. http://dx.doi.org/10.1016/j.ejor.2016.10.044.

Guo, C., Abbasi Shureshjani, R., Foroughi, A. A., & Zhu, J. (2017). Decomposition weights and overall efficiency in two-stage additive network DEA. *European Journal of Operational Research, 257*(3), 896–906. http://dx.doi.org/10.1016/j.ejor.2016.08.002.

Herrera-Restrepo, O., Triantis, K., Seaver, W. L., Paradi, J. C., & Zhu, H. (2016). Bank branch operational performance: A robust multivariate and clustering approach. *Expert Systems with Applications, 50*, 107–119. http://dx.doi.org/10.1016/j.eswa.2015.12.025.

Hosseinzadeh Lotfi, F., Jahanshahloo, G. R., Hemati, S., & Givehchi, S. (2012). A new bargaining game model for measuring performance of two-stagenetwork structures. *International Journal of Research in Industrial Engineering, 1*(2), 27–39.

Huang, T.-H., Chen, K.-C., & Lin, C.-I. (2017). An extension from network DEA to copula-based network SFA: Evidence from the U.S. commercial banks in 2009. *The Quarterly Review of Economics and Finance*. http://dx.doi.org/10.1016/j.qref.2017.04.007.

Huang, T.-H., Lin, C.-I., & Chen, K.-C. (2017). Evaluating efficiencies of Chinese commercial banks in the context of stochastic multistage technologies. *Pacific-Basin Finance Journal, 41*, 93–110. http://dx.doi.org/10.1016/j.pacfin.2016.12.008.

Izadikhah, M., & Farzipoor Saen, R. (2016). Evaluating sustainability of supply chains by two-stage range directional measure in the presence of negative data. *Transportation Research Part D: Transport and Environment, 49*, 110–126. http://dx.doi.org/10.1016/j.trd.2016.09.003.

Kamarudin, F., Sufian, F., Loong, F. W., & Anwar, N. A. M. (2017). Assessing the domestic and foreign Islamic banks efficiency: Insights from selected Southeast Asian countries. *Future Business Journal, 3*(1), 33–46. http://dx.doi.org/10.1016/j.fbj.2017.01.005.

Kao, C. (2014). Efficiency decomposition for general multi-stage systems in data envelopment analysis. *European Journal of Operational Research, 232*(1), 117–124. http://dx.doi.org/10.1016/j.ejor.2013.07.012.

Kao, C., & Hwang, S.-N. (2008). Efficiency decomposition in two-stage data envelopment analysis: An application to non-life insurance companies in Taiwan. *European Journal of Operational Research, 185*(1), 418–429. http://dx.doi.org/10.1016/j.ejor.2006.11.041.

Kao, C., & Hwang, S.-N. (2011). Decomposition of technical and scale efficiencies in two-stage production systems. *European Journal of Operational Research, 211*(3), 515–519. http://dx.doi.org/10.1016/j.ejor.2011.01.010.

Kao, C., & Hwang, S.-N. (2014). Multi-period efficiency and Malmquist productivity index in two-stage production systems. *European Journal of Operational Research, 232*(3), 512–521. http://dx.doi.org/10.1016/j.ejor.2013.07.030.

Kwon, H.-B., & Lee, J. (2015). Two-stage production modeling of large U.S. banks: A DEA-neural network approach. *Expert Systems with Applications, 42*(19), 6758–6766. http://dx.doi.org/10.1016/j.eswa.2015.04.062.

Lewis, H. F., Mallikarjun, S., & Sexton, T. R. (2013). Unoriented two-stage DEA: The case of the oscillating intermediate products. *European Journal of Operational Research, 229*(2), 529–539. http://dx.doi.org/10.1016/j.ejor.2013.02.058.

Li, X. (2017). A fair evaluation of certain stage in a two-stage structure: Revisiting the typical two-stage DEA approaches. *Omega, 68*, 155–167. http://dx.doi.org/10.1016/j.omega.2016.07.002.

Li, Y., Chen, Y., Liang, L., & Xie, J. (2012). DEA models for extended two-stage network structures. *Omega, 40*(5), 611–618. http://dx.doi.org/10.1016/j.omega.2011.11.007.

Li, Y., & Cui, Q. (2017). Carbon neutral growth from 2020 strategy and airline environmental inefficiency: A network range adjusted environmental data envelopment analysis. *Applied Energy, 199*, 13–24. http://dx.doi.org/10.1016/j.apenergy.2017.04.072.

Liang, L., Cook, W. D., & Zhu, J. (2008). DEA models for two-stage processes: Game approach and efficiency decomposition. *Naval Research Logistics (NRL), 55*(7), 643–653. doi:10.1002/nav.20308.

Liang, L., Yang, F., Cook, W. D., & Zhu, J. (2006). DEA models for supply chain efficiency evaluation. *Annals of Operations Research, 145*, 35–49.

Liu, S.-T. (2011a). A note on efficiency decomposition in two-stage data envelopment analysis. *European Journal of Operational Research, 212*(3), 606–608. http://dx.doi.org/10.1016/j.ejor.2011.03.001.

Liu, S.-T. (2011b). Performance measurement of Taiwan financial holding companies: An additive efficiency decomposition approach. *Expert Systems with Applications, 38*(5), 5674–5679. http://dx.doi.org/10.1016/j.eswa.2010.10.068.

Liu, S.-T. (2014a). Fuzzy efficiency ranking in fuzzy two-stage data envelopment analysis. *Optimization Letters, 8*(2), 633–652. doi:10.1007/s11590-012-0602-5.

Liu, S.-T. (2014b). Restricting weight flexibility in fuzzy two-stage DEA. *Computers & Industrial Engineering, 74*(0), 149–160. http://dx.doi.org/10.1016/j.cie.2014.05.011.

Liu, W., Zhou, Z., Ma, C., Liu, D., & Shen, W. (2015). Two-stage DEA models with undesirable input-intermediate-outputs. *Omega, 56*(0), 74–87. http://dx.doi.org/10.1016/j.omega.2015.03.009.

Lovell, C. A. K., Walters, L. C., & Wood, L. L. (1994). Stratified models of education production using modified DEA and regression analysis. In A. Charnes, W. W. Cooper, A. Y. Lewin, & L. M. Seiford (Eds.), *Data envelopment analysis: Theory, methodology, and application* (pp. 329–351). Boston: Kluwer.

Luo, X. (2003). Evaluating the profitability and marketability efficiency of large banks: An application of data envelopment analysis. *Journal of Business Research, 56*(8), 627–635. http://dx.doi.org/10.1016/S0148-2963(01)00293-4.

Ma, J. (2015). A two-stage DEA model considering shared inputs and free intermediate measures. *Expert Systems with Applications, 42*(9), 4339–4347. http://dx.doi.org/10.1016/j.eswa.2015.01.040.

Mahdiloo, M., Jafarzadeh, A. H., Saen, R. F., Tatham, P., & Fisher, R. (2016). A multiple criteria approach to two-stage data envelopment analysis. *Transportation Research Part D: Transport and Environment, 46*, 317–327. http://dx.doi.org/10.1016/j.trd.2016.04.008.

Matthews, K. (2013). Risk management and managerial efficiency in Chinese banks: A network DEA framework. *Omega, 41*(2), 207–215. http://dx.doi.org/10.1016/j.omega.2012.06.003.

Menon, K., & O'Connor, A. (2007). Building customers' affective commitment towards retail banks: The role of CRM in each 'moment of truth'. *Journal of Financial Services Marketing, 12*(2), 157–168. doi:10.1057/palgrave.fsm.4760068.

Noulas, A. G., Hatzigayios, T., Lazaridis, J., & Lyroudi, K. (2001). Non-parametric production frontier approach to the study of efficiency of non-life insurance companies in Greece. *Journal of Financial Management and Analysis: International Review of Finance, 14*(1), 19–26.

Ohsato, S., & Takahashi, M. (2015). Management efficiency in Japanese regional banks: A network DEA. *Procedia - Social and Behavioral Sciences, 172*, 511–518. http://dx.doi.org/10.1016/j.sbspro.2015.01.394.

Ray, S. (2016). Cost efficiency in an Indian bank branch network: A centralized resource allocation model. *Omega, 65*, 69–81. http://dx.doi.org/10.1016/j.omega.2015.12.009.

Rho, S., & An, J. (2007). Evaluating the efficiency of a two-stage production process using data envelopment analysis. *International Transactions in Operational Research, 14*, 395–410. doi:10.1111/j.1475-3995.2007.00597.x.

Sahin, G., Gokdemir, L., & Ozturk, D. (2016). Global crisis and its effect on Turkish banking sector: A study with data envelopment analysis. *Procedia Economics and Finance, 38*, 38–48. http://dx.doi.org/10.1016/S2212-5671(16)30174-5.

Sahoo, B. K., Zhu, J., Tone, K., & Klemen, B. M. (2014). Decomposing technical efficiency and scale elasticity in two-stage network DEA. *European Journal of Operational Research, 233*(3), 584–594. http://dx.doi.org/10.1016/j.ejor.2013.09.046.

San-Jose, L., Retolaza, J. L., & Torres Pruñonosa, J. (2014). Efficiency in Spanish banking: A multistakeholder approach analysis. *Journal of International Financial Markets, Institutions and Money, 32*, 240–255. http://dx.doi.org/10.1016/j.intfin.2014.06.005.

Schinnar, A. P., Kamis-Gould, E., Delucia, N., & Rothbard, A. B. (1990). Organizational determinants of efficiency and effectiveness in mental health partial care programs. *Health Services Research, 25*, 387–420.

Seiford, L. M., & Zhu, J. (1999). Profitability and marketability of the Top 55 U.S. commercial banks. *Management Science, 45*(9), 1270–1288. doi:10.1287/mnsc.45.9.1270.

Sexton, T., & Lewis, H. (2003). Two-stage DEA: An application to major league baseball. *Journal of Productivity Analysis, 19*(2-3), 227–249. doi:10.1023/A:1022861618317.

Silva, T. C., Tabak, B. M., Cajueiro, D. O., & Dias, M. V. B. (2017). A comparison of DEA and SFA using micro- and macro-level perspectives: Efficiency of Chinese local banks. *Physica A: Statistical Mechanics and its Applications, 469*, 216–223. http://dx.doi.org/10.1016/j.physa.2016.11.041.

Stewart, C., Matousek, R., & Nguyen, T. N. (2016). Efficiency in the Vietnamese banking system: A DEA double bootstrap approach. *Research in International Business and Finance, 36*, 96–111. http://dx.doi.org/10.1016/j.ribaf.2015.09.006.

Sun, B.-Q., Xiong, S., & Tang, K. (2013). A modification of relational two-stage DEA with variable returns to scale. *Journal of Software, 8*(8), 1859–1866.

Tavana, M., & Khalili-Damghani, K. (2014). A new two-stage Stackelberg fuzzy data envelopment analysis model. *Measurement, 53*(0), 277–296. http://dx.doi.org/10.1016/j.measurement.2014.03.030.

Toloo, M. (2014). The role of non-Archimedean epsilon in finding the most efficient unit: With an application of professional tennis players. *Applied Mathematical Modelling, 38*, 5334–5346.

Toloo, M., Emrouznejad, A., & Moreno, P. (2015). A linear relational DEA model to evaluate two-stage processes with shared inputs. *Computational and Applied Mathematics*, 1–17. doi:10.1007/s40314-014-0211-2.

Toloo, M., & Tavana, M. (2017). A novel method for selecting a single efficient unit in data envelopment analysis without explicit inputs/outputs. *Annals of Operations Research, 253*, 657–681.

Wang, C., Gopal, R., & Zionts, S. (1997). Use of data envelopment analysis in assessing information technology impact on firm performance. *Annals of Operations Research, 73*(0), 191–213. doi:10.1023/A:1018977111455.

Wang, K., Huang, W., Wu, J., & Liu, Y.-N. (2014). Efficiency measures of the Chinese commercial banking system using an additive two-stage DEA. *Omega, 44*, 5–20. http://dx.doi.org/10.1016/j.omega.2013.09.005.

Wang, Q., Hang, Y., Sun, L., & Zhao, Z. (2016). Two-stage innovation efficiency of new energy enterprises in China: A non-radial DEA approach. *Technological Forecasting and Social Change, 112*, 254–261. http://dx.doi.org/10.1016/j.techfore.2016.04.019.

Wang, Y.-M., & Chin, K.-S. (2010). Some alternative DEA models for two-stage process. *Expert Systems with Applications, 37*(12), 8799–8808. http://dx.doi.org/10.1016/j.eswa.2010.06.024.

Wanke, P., & Barros, C. (2014). Two-stage DEA: An application to major Brazilian banks. *Expert Systems with Applications, 41*(5), 2337–2344. http://dx.doi.org/10.1016/j.eswa.2013.09.031.

Wanke, P., Barros, C. P., & Emrouznejad, A. (2016). Assessing productive efficiency of banks using integrated Fuzzy-DEA and bootstrapping: A case of Mozambican banks. *European Journal of Operational Research, 249*(1), 378–389. http://dx.doi.org/10.1016/j.ejor.2015.10.018.

Wanke, P., Maredza, A., & Gupta, R. (2017). Merger and acquisitions in South African banking: A network DEA model. *Research in International Business and Finance, 41*, 362–376. http://dx.doi.org/10.1016/j.ribaf.2017.04.055.

Widiarto, I., Emrouznejad, A., & Anastasakis, L. (2017). Observing choice of loan methods in not-for-profit microfinance using data envelopment analysis. *Expert Systems with Applications, 82*, 278–290. http://dx.doi.org/10.1016/j.eswa.2017.03.022.

Wu, J., Yin, P., Sun, J., Chu, J., & Liang, L. (2016). Evaluating the environmental efficiency of a two-stage system with undesired outputs by a DEA approach: An interest preference perspective. *European Journal of Operational Research, 254*(3), 1047–1062. http://dx.doi.org/10.1016/j.ejor.2016.04.034.

Yang, Z. (2006). A two-stage DEA model to evaluate the overall performance of Canadian life and health insurance companies. *Mathematical and Computer Modelling, 43*(7–8), 910–919. http://dx.doi.org/10.1016/j.mcm.2005.12.011.

Yu, Y., Shi, Q.-F., & Song, J. (2013). A note on some alternative DEA models for two-stage process. *Expert Systems with Applications, 40*(10), 4268–4269. http://dx.doi.org/10.1016/j.eswa.2013.01.022.

Yu, Y., & Shi, Q. (2014). Two-stage DEA model with additional input in the second stage and part of intermediate products as final output. *Expert Systems with Applications, 41*(15), 6570–6574. http://dx.doi.org/10.1016/j.eswa.2014.05.021.

Zha, Y., & Liang, L. (2010). Two-stage cooperation model with input freely distributed among the stages. *European Journal of Operational Research, 205*(2), 332–338. http://dx.doi.org/10.1016/j.ejor.2010.01.010.

Zhou, Z., Lin, L., Xiao, H., Ma, C., & Wu, S. (2017). Stochastic network DEA models for two-stage systems under the centralized control organization mechanism. *Computers & Industrial Engineering, 110*, 404–412. https://doi.org/10.1016/j.cie.2017.06.005.

Zhu, J. (2000). Multi-factor performance measure model with an application to Fortune 500 companies. *European Journal of Operational Research, 123*(1), 105–124. http://dx.doi.org/10.1016/S0377-2217(99)00096-X.