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

Mohammad Izadikhah\textsuperscript{a}, Madjid Tavana\textsuperscript{b,c,*}, Debora Di Caprio\textsuperscript{d,e}, Francisco J. Santos-Arteaga\textsuperscript{f}

\textsuperscript{a} Department of Mathematics, College of Science, Arak Branch, Islamic Azad University, Arak, Iran
\textsuperscript{b} Business Systems and Analytics Department, Distinguished Chair of Business Analytics, La Salle University, Philadelphia, PA 19141, USA
\textsuperscript{c} Business Information Systems Department, Faculty of Business Administration and Economics, University of Paderborn, D-33098 Paderborn, Germany
\textsuperscript{d} Department of Mathematics and Statistics, York University, Toronto, M3J 1P3, Canada
\textsuperscript{e} Polo Tecnologico IES G. Galilei, Via Cadorna 14, 21300, Bolzano, Italy
\textsuperscript{f} Faculty of Economics and Management, Free University of Bolzano, Bolzano, Italy

A R T I C L E   I N F O

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

A B S T R A C T

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.

© 2017 Elsevier Ltd. All rights reserved.

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-
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, Pentsi, 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, & Givelchli, 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 of 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, Copal, and Zions (1997) and Noulas, Hatzigiovi, Lazaridis, and Lyroudi (2001) proposed applications of two-stage DEA to information technology (IT) and to non-life insurance policies, respectively. Färe and Grosskopf (1996) 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 non-linear 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. Fukuyma 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. Akter 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 Khalilli-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 Malquist 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.

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), Fukuyma 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 keynotes models in the literature that will be extended by our approach.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( DMU_{j} )</td>
<td>First stage DMU</td>
</tr>
<tr>
<td>( DMU_{p} )</td>
<td>Second stage DMU</td>
</tr>
<tr>
<td>( x_{i,j} )</td>
<td>Input of the first stage</td>
</tr>
<tr>
<td>( z_{d,j} )</td>
<td>Output of the first stage</td>
</tr>
<tr>
<td>( y_{r,p} )</td>
<td>Output of the second stage</td>
</tr>
<tr>
<td>( s_{p} )</td>
<td>Overall efficiency of the generic ( p )th DMU, ( DMU_{p} )</td>
</tr>
<tr>
<td>( s_{1,p} )</td>
<td>Efficiency of Stage 1 (the first sub-DMU) of ( DMU_{p} )</td>
</tr>
<tr>
<td>( s_{2,p} )</td>
<td>Efficiency of Stage 2 (the second sub-DMU) of ( DMU_{p} )</td>
</tr>
<tr>
<td>( s_{1} )</td>
<td>Lower bound of the efficiency of Stage 1 of ( DMU_{p} )</td>
</tr>
<tr>
<td>( s_{2} )</td>
<td>Lower bound of the efficiency of Stage 2 of ( DMU_{p} )</td>
</tr>
<tr>
<td>( \varepsilon )</td>
<td>Non-Archimedean positive small number</td>
</tr>
</tbody>
</table>

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 decompositions 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 \( p \)-th DMU, \( DMU_p \):

\[
\theta_p = \max \sum_{d=1}^{D} w_d z_{dp}
\]

s.t.

\[
\begin{align*}
\sum_{i=1}^{m} v_i x_{ip} &= 1, \\
\sum_{i=1}^{5} u_i y_{ij} - \sum_{i=1}^{m} v_i x_{ij} &\leq 0, \quad \forall j = 1, \ldots, n; \\
\frac{D}{d} \sum_{d=1}^{D} w_d z_{dj} - \sum_{i=1}^{m} v_i x_{ij} &\leq 0, \quad \forall j = 1, \ldots, n; \\
\sum_{i=1}^{5} u_i y_{ij} - \sum_{i=1}^{m} w_d z_{dj} &\leq 0, \quad \forall j = 1, \ldots, n; \\
u_r, u_i, w_d &\geq \varepsilon, \quad \forall r = 1, \ldots, s; \forall i = 1, \ldots, m; \forall d = 1, \ldots, D.
\end{align*}
\]

(1)

They introduced the non-Archimedean positive small number, \( \varepsilon \), 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 \( \varepsilon \) 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:

\[
\theta_{1p} = \max \sum_{d=1}^{D} w_d z_{dp}
\]

s.t.

\[
\begin{align*}
\sum_{i=1}^{m} v_i x_{ip} &= 1, \\
\sum_{i=1}^{5} u_i y_{ip} - \theta_p \sum_{i=1}^{m} v_i x_{ip} &= 0, \\
\sum_{i=1}^{5} u_i y_{ij} - \sum_{i=1}^{m} v_i x_{ij} &\leq 0, \quad \forall j = 1, \ldots, n; \\
\frac{D}{d} \sum_{d=1}^{D} w_d z_{dj} - \sum_{i=1}^{m} v_i x_{ij} &\leq 0, \quad \forall j = 1, \ldots, n; \\
\sum_{i=1}^{5} u_i y_{ij} - \sum_{i=1}^{m} w_d z_{dj} &\leq 0, \quad \forall j = 1, \ldots, n; \\
u_r, u_i, w_d &\geq \varepsilon, \quad \forall r = 1, \ldots, s; \forall i = 1, \ldots, m; \forall d = 1, \ldots, D.
\end{align*}
\]

(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 \( \alpha_i \) is assigned to each input quantity \( x_{ij} \) of a given \( DMU_i \). The factor \( \alpha_i \) 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 the two stages of the production process.

\[
\theta_p = \max \sum_{d=1}^{D} w_d z_{dp}
\]

s.t.

\[
\begin{align*}
\sum_{i=1}^{m} v_i x_{ip} &= 1, \\
\sum_{i=1}^{5} u_i y_{ip} - \theta_p \sum_{i=1}^{m} v_i x_{ip} &= 0, \\
\sum_{i=1}^{5} u_i y_{ij} - \sum_{i=1}^{m} v_i x_{ij} &\leq 0, \quad \forall j = 1, \ldots, n; \\
\frac{D}{d} \sum_{d=1}^{D} w_d z_{dj} - \sum_{i=1}^{m} v_i x_{ij} &\leq 0, \quad \forall j = 1, \ldots, n; \\
\sum_{i=1}^{5} u_i y_{ij} - \sum_{i=1}^{m} w_d z_{dj} &\leq 0, \quad \forall j = 1, \ldots, n; \\
u_r, u_i, w_d &\geq \varepsilon, \quad \forall r = 1, \ldots, s; \forall i = 1, \ldots, m; \forall d = 1, \ldots, D.
\end{align*}
\]

(3)

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

Then, they applied a heuristic method to transform the above nonlinear programming model into a parametric linear one, where \( \delta \) 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 \( \alpha_d \) 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, \( \phi_d = w_d \alpha_d \) \( (d = 1, 2, \ldots, D) \) denotes the weight of the portion of the \( d \)th intermediate 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, \( \text{DMU}_p \), 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).

\[
\theta_{1p}^{u} = \max \sum_{d=1}^{D} w_d z_{dp} \\
\text{s.t.} \sum_{i=1}^{m} v_i x_{ip} = 1, \\
\sum_{d=1}^{D} w_d z_{dj} - \sum_{i=1}^{m} v_i x_{ij} \leq 0, \quad \forall j = 1, \ldots, n; \\
w_d, v_i \geq 0, \quad \forall d = 1, \ldots, D; \ i = 1, \ldots, m.
\]

and

\[
\theta_{2p}^{l} = \max \sum_{r=1}^{s} u_r y_{rp} \\
\text{s.t.} \sum_{d=1}^{D} \phi_d z_{dp} + \sum_{h=1}^{H} q_h x_{hp}^{2} = 1, \\
\sum_{r=1}^{s} u_r y_{rj} - \sum_{d=1}^{D} \phi_d z_{dj} - \sum_{h=1}^{H} q_h x_{hj}^{2} \leq 0, \quad \forall j = 1, \ldots, n; \\
u_r, \phi_d, q_h, w_d, v_i \geq 0, \quad \forall r=1, \ldots, s; \ d=1, \ldots, D; \ h=1, \ldots, H; \ i=1, \ldots, m.
\]

**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.

\[
\theta_{2p}^{u} = \max \sum_{r=1}^{s} u_r y_{rp} \\
\text{s.t.} \sum_{d=1}^{D} \phi_d z_{dp} + \sum_{h=1}^{H} q_h x_{hp}^{2} = 1, \\
\sum_{r=1}^{s} u_r y_{rj} - \sum_{d=1}^{D} \phi_d z_{dj} - \sum_{h=1}^{H} q_h x_{hj}^{2} \leq 0, \quad \forall j = 1, \ldots, n; \\
u_r, \phi_d, q_h, w_d, v_i \geq 0, \quad \forall r=1, \ldots, s; \ d=1, \ldots, D; \ h=1, \ldots, H; \ i=1, \ldots, m.
\]

Hence, Yu and Shi (2014) focused on the cooperative efficiency of the generic \( p \)-th DMU, \( \text{DMU}_p \). A nonlinear fractional programming problem was used to model the overall efficiency of \( \text{DMU}_p \) and the Charnes–Cooper transformation applied to restate it as follows:

\[
\theta_p = \max \sum_{d=1}^{D} \phi_d z_{dp} \\
\text{s.t.} \sum_{i=1}^{m} v_i x_{ip} = 1, \\
\sum_{d=1}^{D} \phi_d z_{dp} + \sum_{h=1}^{H} q_h x_{hp}^{2} = 1, \\
\sum_{r=1}^{s} u_r y_{rj} - \sum_{d=1}^{D} \phi_d z_{dj} - \sum_{h=1}^{H} q_h x_{hj}^{2} \leq 0, \quad \forall j = 1, \ldots, n; \\
u_r, \phi_d, q_h, w_d, v_i \geq 0, \quad \forall r=1, \ldots, s; \ d=1, \ldots, D; \ h=1, \ldots, H; \ i=1, \ldots, m.
\]

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

$$\theta_p = \max \lambda \times \sum_{i=1}^{m} u_i y_{rp}$$

s.t.

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

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

$$\sum_{d=1}^{D} w_d z_{dj} \leq \sum_{i=1}^{m} u_i x_{ij}, \quad \forall j = 1, ..., n;$$

$$\sum_{i=1}^{m} u_i y_{ij} \leq \sum_{d=1}^{D} \varphi_d z_{dp} + \sum_{h=1}^{H} q_h x_{hp}^2, \quad \forall j = 1, ..., n;$$

$$w_d \geq \varphi_d \geq 0, \quad \forall d = 1, ..., D; \quad \forall r = 1, ..., s; \quad d = 1, ..., D; \quad h = 1, ..., H; \quad i = 1, ..., m.$$  \hspace{1cm} (9)

where $\lambda \in [0, 1]$ can be treated as a parameter. An alternative transformation of Model (8) can be carried out by setting $\lambda = \sum_{i=1}^{m} u_i y_{rp}$ and $\theta_p^{U} \leq \lambda \leq \theta_p^{L}$.

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 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:

- $x_{ij}, i = 1, ..., I$ Inputs consumed by Stage 1 of DMU$_j$ entirely.
- $x_{ij}^2, j = 1, ..., T$ Inputs freely allocated to both sub-DMUs of DMU$_j$.
- $x_{ij}^3, k = 1, ..., K$ Inputs consumed by Stage 2 of DMU$_j$ entirely.
- $z_{ij}^2, d = 1, ..., D$ Intermediate outputs consumed by Stage 2 of DMU$_j$ entirely.
- $z_{ij}^3, b = 1, ..., B$ Intermediate outputs acting as inputs of Stage 2 and final outputs of DMU$_j$.
- $y_{ij}^1, h = 1, ..., H$ Final outputs produced by Stage 1 of DMU$_j$.
- $y_{ij}^2, r = 1, ..., R$ Final outputs produced by Stage 2 of DMU$_j$.

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

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

$$\alpha_t, t = 1, ..., T$$

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

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

$$\beta_b, b = 1, ..., B$$

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

- $\beta^2 z_{ij}^2$ is the input of the sub-DMU in Stage 2.
- $(1-\beta^2)z_{ij}^2$ 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$_p$.

$$\theta_{1p}^{U} = \max \sum_{d=1}^{D} w_{dp}^1 + \sum_{b=1}^{B} w_{bp}^2 + \sum_{h=1}^{H} u_{hp}^1 \quad \text{s.t.} \quad \sum_{i=1}^{m} v_{i} x_{ip}^1 + \sum_{t=1}^{T} v_{t} (\alpha x_{ip}^2) \leq 1, \quad \forall j = 1, ..., n,$$

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

$$0 \leq \alpha \leq 1.$$

(10)

By using the Charnes–Cooper transformation Charnes & Cooper, 1962) and setting $\mu_t^2 = v_t^2 \alpha_t (t = 1, ..., T)$, Model (10) can be transformed into the following linear model:
\( \theta_{1p}^{11} = \max \sum_{d=1}^{D} w_d^1 x_{dp} + \sum_{b=1}^{B} w_b^1 x_{bp}^2 + \sum_{h=1}^{H} u_h^1 y_{hp} \)

s.t. 
\( \sum_{t=1}^{I} v_t^1 x_{1tp} + \sum_{t=1}^{T} \mu_t^2 x_{2tp} = 1. \)
\( D \sum_{d=1}^{D} w_d^2 x_{1d}^1 + \sum_{b=1}^{B} w_b^2 x_{2b}^1 + \sum_{h=1}^{H} u_h^2 y_{h1}^1 - \sum_{t=1}^{I} v_t^1 x_{1t}^1 - \sum_{t=1}^{T} \mu_t^2 x_{2t}^1 \leq 0. \)
\( \forall j = 1, ..., n. \)
\( w_d^1, w_b^1, u_h^1, v_t^1, \mu_t^2 \geq 0. \)
\( \forall d = 1, ..., D; \ b = 1, ..., B; \ h = 1, ..., H; \ i = 1, ..., m; \ t = 1, ..., T. \)

\[ (11) \]

**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\(_p\).

\( \theta_{2p}^i = \max \sum_{k=1}^{R} u_k^i y_{kp}^2 \)

s.t. 
\( \sum_{k=1}^{R} u_k^i y_{kp}^2 \leq 1. \)
\( \forall j = 1, ..., n. \)
\( \sum_{k=1}^{R} v_k^i y_{1kp}^1 + \sum_{t=1}^{T} (1-\alpha_t) y_{2kp}^1 + \sum_{d=1}^{D} w_d^i x_{1dp} + \sum_{b=1}^{B} w_b^i x_{2bp} \)
\( = \theta_{1p}^{11}. \)
\( \forall j = 1, ..., n. \)
\( \sum_{k=1}^{R} v_k^i y_{1kp}^1 + \sum_{t=1}^{T} (1-\alpha_t) y_{2kp}^1 + \sum_{d=1}^{D} w_d^i x_{1dp} + \sum_{b=1}^{B} w_b^i x_{2bp} \)
\( \leq 1. \)
\( \forall j = 1, ..., n. \)
\( w_d^i, w_b^i, u_h^i, v_t^1, \mu_t^2 \geq 0. \)
\( \forall d = 1, ..., D; \ b = 1, ..., B; \ h = 1, ..., H; \ i = 1, ..., I; \ t = 1, ..., T. \)

\[ (12) \]

By using the Charnes–Cooper transformation (Charnes & Cooper, 1962) and setting \( \mu_t^2 = v_t^1 (1 - \alpha_t) \) and \( \psi_h^2 = w_h^2 b_h \), Model (12) can be transformed into the following linear model:

\( \theta_{2p}^i = \max \sum_{k=1}^{R} u_k^i y_{kp}^2 \)

s.t. 
\( \sum_{k=1}^{K} v_k^i y_{1kp}^1 + \sum_{t=1}^{T} (\mu_t^2 - v_t^1) y_{2kp}^1 + \sum_{d=1}^{D} w_d^i x_{1dp} + \sum_{b=1}^{B} w_b^i x_{2bp} = 1. \)
\( \forall j = 1, ..., n. \)
\( w_d^i, u_h^i, v_t^1, \mu_t^2 \geq 0. \)
\( \forall d = 1, ..., D; \ b = 1, ..., B; \ m = 1, ..., M; \ r = 1, ..., R; \ k = 1, ..., K; \ t = 1, ..., T. \)

\[ (13) \]

**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\(_p\).

\( \theta_{1p}^{11} = \max \sum_{d=1}^{D} w_d^1 x_{1dp} + \sum_{b=1}^{B} w_b^1 x_{2bp} + \sum_{h=1}^{H} u_h^1 y_{hp} \)

s.t. 
\( \sum_{d=1}^{D} w_d^2 x_{1d}^1 + \sum_{b=1}^{B} w_b^2 x_{2b}^1 + \sum_{h=1}^{H} u_h^2 y_{h1}^1 - \sum_{t=1}^{T} v_t^1 x_{1t}^1 - \sum_{t=1}^{T} \mu_t^2 x_{2t}^1 \leq 0. \)
\( \forall j = 1, ..., n. \)
\( w_d^1, w_b^1, u_h^1, v_t^1, \mu_t^2 \geq 0. \)
\( \forall d = 1, ..., D; \ b = 1, ..., B; \ r = 1, ..., s; \ k = 1, ..., K; \ h = 1, ..., H; \ i = 1, ..., I; \ t = 1, ..., T. \)

\[ (13) \]

\[ (14) \]

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

\( \theta_{1p}^{11} = \max \sum_{d=1}^{D} w_d^1 x_{1dp} + \sum_{b=1}^{B} w_b^1 x_{2bp} + \sum_{h=1}^{H} u_h^1 y_{hp} \)

s.t. 
\( \sum_{d=1}^{D} w_d^2 x_{1d}^1 + \sum_{b=1}^{B} w_b^2 x_{2b}^1 + \sum_{h=1}^{H} u_h^2 y_{h1}^1 - \sum_{t=1}^{T} v_t^1 x_{1t}^1 - \sum_{t=1}^{T} \mu_t^2 x_{2t}^1 \leq 0. \)
\( \forall j = 1, ..., n. \)
\( w_d^1, w_b^1, u_h^1, v_t^1, \mu_t^2 \geq 0. \)
\( \forall d = 1, ..., D; \ b = 1, ..., B; \ r = 1, ..., s; \ k = 1, ..., K; \ h = 1, ..., H; \ i = 1, ..., I; \ t = 1, ..., T. \)

\[ (15) \]

**The follower’s efficiency:**

Model (16) evaluates the efficiency of the sub-DMU in Stage 2 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}^I = \max \sum_{d=1}^D \sum_{b=1}^B w_d^{x_{dp}^1} + \sum_{b=1}^B w_b^{x_{bp}^{221}} + \sum_{h=1}^H u_h^{y_{hp}^{221}}$

s.t.

$$\sum_{i=1}^I v_i^j x_{ip}^1 + \sum_{t=1}^T v_t^j (\alpha x_{tp}^1) \leq 1, \quad \forall j = 1, \ldots, n.$$

$$\sum_{k=1}^K v_k^{x_{kp}^2} + \sum_{t=1}^T v_t^2 (1 - \alpha_i) x_{tp}^2 + \sum_{d=1}^D w_d^{x_{dp}^1} + \sum_{b=1}^B w_b^{x_{bp}^{221}} \leq \beta_b^{x_{bp}^{221}},$$

$$\sum_{k=1}^K v_k^{x_{kp}^2} + \sum_{t=1}^T v_t^2 (1 - \alpha_i) x_{tp}^2 + \sum_{d=1}^D w_d^{x_{dp}^1} + \sum_{b=1}^B w_b^{x_{bp}^{221}} \leq 1, \quad \forall j = 1, \ldots, n.$$

$$w_d^{x_{dp}^1}, w_b^{x_{bp}^{221}}, u_k^{x_{kp}^2}, v_t^{x_{tp}^2}, v_t^j, u_h^y \geq 0. \quad \forall d = 1, \ldots, D; b = 1, \ldots, B; r = 1, \ldots; s; k = 1, \ldots; \mu_b \leq \beta_b \leq 1,$$

$$\theta_{1p} = \max \sum_{d=1}^D \sum_{b=1}^B w_d^{x_{dp}^1} + \sum_{b=1}^B w_b^{x_{bp}^{221}} + \sum_{h=1}^H u_h^{y_{hp}^{221}}$$

s.t.

$$\sum_{i=1}^I v_i^j x_{ip}^1 + \sum_{t=1}^T v_t^j (\alpha x_{tp}^1) \leq 1, \quad \forall j = 1, \ldots, n.$$

$$\sum_{k=1}^K v_k^{x_{kp}^2} + \sum_{t=1}^T v_t^2 (1 - \alpha_i) x_{tp}^2 + \sum_{d=1}^D w_d^{x_{dp}^1} + \sum_{b=1}^B w_b^{x_{bp}^{221}} \leq 1, \quad \forall j = 1, \ldots, n.$$

$$w_d^{x_{dp}^1}, w_b^{x_{bp}^{221}}, u_k^{x_{kp}^2}, v_t^{x_{tp}^2}, v_t^j, u_h^y \geq 0. \quad \forall d = 1, \ldots, D; b = 1, \ldots, B; r = 1, \ldots; s; k = 1, \ldots; \mu_b \leq \beta_b \leq 1,$$

$$0 \leq \alpha \leq 1. \quad \forall t = 1, \ldots, T.$$

$$0 \leq \beta_b \leq 1. \quad \forall b = 1, \ldots, B.$$

(16)

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 $\theta_p$ of the generic p-th DMU, $DMU_p$.

$$\theta_p = \max \sum_{d=1}^D \sum_{b=1}^B w_d^{x_{dp}^1} + \sum_{b=1}^B w_b^{x_{bp}^{221}} + \sum_{h=1}^H u_h^{y_{hp}^{221}}$$

s.t.

$$\sum_{k=1}^K v_k^{x_{kp}^2} + \sum_{t=1}^T v_t^{x_{tp}^2} (1 - \alpha_i) x_{tp}^2 + \sum_{d=1}^D w_d^{x_{dp}^1} + \sum_{b=1}^B w_b^{x_{bp}^{221}} \leq 1, \quad \forall j = 1, \ldots, n.$$

$$w_d^{x_{dp}^1}, w_b^{x_{bp}^{221}}, u_k^{x_{kp}^2}, v_t^{x_{tp}^2}, v_t^j, u_h^y \geq 0. \quad \forall d = 1, \ldots, D; b = 1, \ldots, B; r = 1, \ldots; s; k = 1, \ldots; \mu_b \leq \beta_b \leq 1,$$

$$0 \leq \alpha \leq 1. \quad \forall t = 1, \ldots, T.$$

$$0 \leq \beta_b \leq 1. \quad \forall b = 1, \ldots, B.$$

(18)

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

$$\theta_p = \max \left( \frac{\sum_{d=1}^D \sum_{b=1}^B w_d^{x_{dp}^1} + \sum_{b=1}^B w_b^{x_{bp}^{221}} + \sum_{h=1}^H u_h^{y_{hp}^{221}}}{R \sum_{r=1}^R u_r^{y_{rp}^2}} \right)$$

s.t.

$$\sum_{i=1}^I v_i^j x_{ip}^1 + \sum_{t=1}^T v_t^j (\alpha x_{tp}^1) = 1, \quad \forall j = 1, \ldots, n.$$

$$\sum_{k=1}^K v_k^{x_{kp}^2} + \sum_{t=1}^T v_t^2 (1 - \alpha_i) x_{tp}^2 + \sum_{d=1}^D w_d^{x_{dp}^1} + \sum_{b=1}^B w_b^{x_{bp}^{221}} = 1,$$

$$w_d^{x_{dp}^1}, w_b^{x_{bp}^{221}}, u_k^{x_{kp}^2}, v_t^{x_{tp}^2}, v_t^j, u_h^y \geq 0. \quad \forall d = 1, \ldots, D; b = 1, \ldots, B; r = 1, \ldots; s; k = 1, \ldots; \mu_b \leq \beta_b \leq 1,$$

$$0 \leq \alpha \leq 1. \quad \forall t = 1, \ldots, T.$$

$$0 \leq \beta_b \leq 1. \quad \forall b = 1, \ldots, B.$$

(17)
\[ k = 1, \ldots, K; t = 1, \ldots, T; i = 1, \ldots, I; h = 1, \ldots, H. \]
\[ 0 \leq \theta^2_b \leq \theta^2_b, \quad \forall b = 1, \ldots, B. \]
\[ 0 \leq \mu^2_t \leq \mu^2_t, \quad \forall t = 1, \ldots, T. \]

(19)

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

\[
\begin{align*}
\theta^1_{tp} &\leq \sum_{d=1}^{D} w^*_{1d} x^1_d + \sum_{b=1}^{B} w^*_{1b} x^1_b + \sum_{h=1}^{H} u^1_{hp} \leq \theta^1_{tp} \quad \text{and} \\
\theta^2_{tp} &\leq \sum_{r=1}^{R} u^2_{rtp} \leq \theta^2_{tp}.
\end{align*}
\]

Therefore, letting \( \lambda = \sum_{d=1}^{D} w^*_{1d} x^1_d + \sum_{b=1}^{B} w^*_{1b} x^1_b + \sum_{h=1}^{H} u^1_{hp} \), Model (19) becomes the following nonlinear programming model:

\[
\begin{align*}
\theta_p &= \max \lambda \times \left( \sum_{r=1}^{R} u^2_{rtp} \right) \\
\text{s.t.,} \\
\sum_{t=1}^{T} \nu^1_{tp} x^1_t + \sum_{t=1}^{T} \mu^2_t x^2_t &= 1, \\
\sum_{k=1}^{K} \nu^3_{kp} x^3_k + \sum_{t=1}^{T} (\nu^2_t - \mu^2_t) x^2_t &= 1, \\
\sum_{d=1}^{D} w^*_{1d} x^1_d + \sum_{b=1}^{B} w^*_{1b} x^1_b + \sum_{h=1}^{H} u^1_{hp} x^1_h - \sum_{i=1}^{I} \nu^1_{ti} x^1_i - \sum_{j=1}^{J} \mu^2_{tj} x^2_j &= 0, \\
\forall j = 1, \ldots, n, \\
\sum_{r=1}^{R} u^2_{rtp} - \sum_{k=1}^{K} \nu^3_{kp} x^3_k + \sum_{t=1}^{T} (\nu^2_t - \mu^2_t) x^2_t - \sum_{d=1}^{D} w^*_{1d} x^1_d + \sum_{b=1}^{B} w^*_{1b} x^1_b + \sum_{h=1}^{H} u^1_{hp} x^1_h - \sum_{i=1}^{I} \nu^1_{ti} x^1_i - \sum_{j=1}^{J} \mu^2_{tj} x^2_j &= 0, \\
\forall j = 1, \ldots, n, \\
w^*_{1d}, w^*_{1b}, u^1_{hp}, v^1_t, v^1_i, u^1_h, \geq 0, \quad \forall d = 1, \ldots, D; b = 1, \ldots, B; r = 1, \ldots, s; \\
k = 1, \ldots, K; t = 1, \ldots, T; i = 1, \ldots, I; h = 1, \ldots, H \\
0 \leq \theta^2_b \leq \theta^2_b, \quad \forall b = 1, \ldots, B. \\
0 \leq \mu^2_t \leq \mu^2_t, \quad \forall t = 1, \ldots, T. \\
\theta^1_{tp} \leq \lambda \leq \theta^1_{tp}
\end{align*}
\]

(20)

Now, consider the following transformations:

\[
\begin{align*}
\tilde{w}^1_d &= \lambda w^1_d, \quad d = 1, \ldots, D, \\
\tilde{w}^2_b &= \lambda w^2_b, \quad b = 1, \ldots, B, \\
\tilde{u}^1_h &= \lambda u^1_h, \quad h = 1, \ldots, H, \\
\tilde{u}^2_i &= \lambda u^2_i, \quad i = 1, \ldots, I, \\
\tilde{v}^1_t &= \lambda v^1_t, \quad t = 1, \ldots, T, \\
\tilde{v}^2_j &= \lambda v^2_j, \quad j = 1, \ldots, J, \\
\tilde{u}^2_b &= \lambda u^2_b, \quad b = 1, \ldots, B, \\
\tilde{u}^2_t &= \lambda u^2_t, \quad t = 1, \ldots, T, \\
\tilde{\mu}^2_t &= \lambda \mu^2_t, \quad t = 1, \ldots, T.
\end{align*}
\]
By using the above transformations, Model (20) can be converted into the following linear programming problem:

\[ \theta_p = \max \sum_{t=1}^{T} \tilde{u}_t^p y_{tp} \]

s.t. 

\[ \sum_{k=1}^{K} \tilde{p}_k^1 x_{tk}^1 + \sum_{t=1}^{T} \tilde{u}_t^2 y_{tp} = \lambda, \]

\[ \sum_{k=1}^{K} \tilde{p}_k^2 x_{tk}^2 + \sum_{t=1}^{T} (\tilde{p}_t^2 - \tilde{u}_t^2) y_{tp} + \sum_{d=1}^{D} \tilde{w}_d^1 x_{tdp}^1 + \sum_{b=1}^{B} \tilde{w}_b^2 z_{btp} = \lambda, \]

\[ \sum_{d=1}^{D} \tilde{w}_d^1 x_{tdj}^1 + \sum_{b=1}^{B} \tilde{w}_b^2 x_{bthj}^1 + \sum_{t=1}^{T} \tilde{p}_t^1 x_{tij}^1 - \sum_{t=1}^{T} \tilde{u}_t^2 x_{tij}^2 \leq 0, \]

\[ \sum_{k=1}^{K} \tilde{p}_k^2 x_{tkj}^2 - \sum_{k=1}^{K} \tilde{p}_k^1 x_{tkj}^1 - \sum_{t=1}^{T} (\tilde{p}_t^2 - \tilde{u}_t^2) x_{tij}^2 - \sum_{d=1}^{D} \tilde{w}_d^2 x_{tdj}^2 - \sum_{b=1}^{B} \tilde{w}_b^2 z_{bthj}^2 \leq 0, \]

\[ \forall j = 1, \ldots, n, \]

\[ \tilde{w}_d^1, \tilde{w}_d^2, \tilde{b}_k^1, \tilde{b}_k^2, \tilde{p}_t^1, \tilde{p}_t^2, \tilde{u}_t^1, \tilde{u}_t^2 \geq 0, \quad \forall d = 1, \ldots, D; \ b = 1, \ldots, B; \ t = 1, \ldots, T; \]

\[ k = 1, \ldots, K; \ i = 1, \ldots, I; \ h = 1, \ldots, H \]

\[ 0 \leq \tilde{Q}_b^2 \leq \tilde{W}_b^2, \quad \forall b = 1, \ldots, B, \]

\[ 0 \leq \tilde{\mu}_t^2 \leq \tilde{\nu}_t^2, \quad \forall t = 1, \ldots, T, \]

\[ \theta_{t1}^p \leq \lambda \leq \theta_{t1}^U, \]

(21)

By solving Model (21), we obtain the optimal value \( \theta_{t1}^p \) of the overall efficiency of the cooperative model. Model (21) also shows that the optimal value of \( \lambda \) coincides with the efficiency score of Stage 1 of DMU\( p \), i.e. \( \theta_{t1}^p = \lambda \). Finally, the efficiency score of Stage 2 of DMU\( p \) can be calculated as \( \theta_{t2}^p = \theta_{t1}^p / \theta_{t1}^U \).

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

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.

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 cannot 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.
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 guarantee 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 ($x_3^1$) representing the total annual operating expenses of the branch. Expenses are also considered in Stage 2. The inputs of Stage 2 are: Employees-2 ($x_2^2$) accounting for the number of employees in Stage 2; and Facilities ($x_3^2$) which are represented as scores between 0 and 60 assigned to the number of ATMs 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 ($z_2^2$), 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 ($y_1^2$), which represent the amount of cash, transfer, and promissory notes; and Net Profits ($y_2^2$) 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>
<thead>
<tr>
<th>Table 3</th>
<th>Numerical comparison among efficiency evaluation methods.</th>
</tr>
</thead>
<tbody>
<tr>
<td>DMU</td>
<td>Stage 1</td>
</tr>
<tr>
<td></td>
<td>KH</td>
</tr>
<tr>
<td>A</td>
<td>0.75</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Table 4</th>
<th>Numerical comparison between the proposed method and Yu and Shi (2014).</th>
</tr>
</thead>
<tbody>
<tr>
<td>DMU</td>
<td>Proposed method</td>
</tr>
<tr>
<td></td>
<td>Overall</td>
</tr>
<tr>
<td>1</td>
<td>0.894</td>
</tr>
<tr>
<td>2</td>
<td>0.612</td>
</tr>
<tr>
<td>3</td>
<td>0.800</td>
</tr>
<tr>
<td>4</td>
<td>0.628</td>
</tr>
<tr>
<td>5</td>
<td>0.408</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>0.833</td>
</tr>
<tr>
<td>8</td>
<td>0.513</td>
</tr>
<tr>
<td>9</td>
<td>0.500</td>
</tr>
<tr>
<td>10</td>
<td>0.596</td>
</tr>
</tbody>
</table>
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 \( \hat{\theta}_2 \) 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 \( \hat{\theta}_2 \) 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

### Table 5
Performance measurement criteria used in the case study.

| 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). |

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

<table>
<thead>
<tr>
<th>Input</th>
<th>Stage 1 input</th>
<th>x₁: Location</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stage 1 input</td>
<td>x₂: Employees-1</td>
</tr>
<tr>
<td></td>
<td>Stage 1 input shared with Stage 2</td>
<td>x₃: Expenses</td>
</tr>
<tr>
<td></td>
<td>Stage 2 input</td>
<td>x₄: Employees-2</td>
</tr>
<tr>
<td></td>
<td>Stage 2 input</td>
<td>x₅: Facilities</td>
</tr>
<tr>
<td>Intermediate</td>
<td>Stage 1 output – Stage 2 input</td>
<td>z₁: Deposits</td>
</tr>
<tr>
<td></td>
<td>Stage 1 output – Stage 2 input</td>
<td>z₂: Loans</td>
</tr>
<tr>
<td></td>
<td>Stage 1 output – Stage 2 input and final output</td>
<td>z₃: Services</td>
</tr>
<tr>
<td>Output</td>
<td>Stage 1 output</td>
<td>y₁: Documents</td>
</tr>
<tr>
<td></td>
<td>Stage 2 output</td>
<td>y₂: Net profits</td>
</tr>
</tbody>
</table>
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 ($Z^2$) 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 block model. This fact is considered by Kao and Hwang (2008) as a signal of inconsistency and inadequacy when evaluating the overall efficiency using the block box model. On the contrary, Xiang and Li (2017) do not see this type of conflicts as a critique to the block 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 8
Descriptive statistics of the two-stage branches of the case study.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>x₁: Location</td>
<td>3.93333</td>
<td>0.707372</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>x₂: Expenses-1</td>
<td>37</td>
<td>5.237229</td>
<td>31</td>
<td>46</td>
</tr>
<tr>
<td>x₂: Expenses-2</td>
<td>6.178,675</td>
<td>4.328,220</td>
<td>2,231,200</td>
<td>16,406,000</td>
</tr>
<tr>
<td>x₃: Facilities</td>
<td>34.93333</td>
<td>13.84334</td>
<td>16</td>
<td>63</td>
</tr>
<tr>
<td>x₄: Deposits</td>
<td>47.13333</td>
<td>7.567474</td>
<td>34</td>
<td>58</td>
</tr>
<tr>
<td>x₅: Loans</td>
<td>32,168,052</td>
<td>29,884,038</td>
<td>9,930,070</td>
<td>1,33E+08</td>
</tr>
<tr>
<td>x₆: Services</td>
<td>2940.667</td>
<td>1071.356</td>
<td>1286</td>
<td>5051</td>
</tr>
<tr>
<td>y₁: Documents</td>
<td>220,382.2</td>
<td>63,701.57</td>
<td>130,134</td>
<td>334,380</td>
</tr>
<tr>
<td>y₂: Net profits</td>
<td>5,727,955</td>
<td>6,934,316</td>
<td>102,780</td>
<td>19,708,980</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>DMU</th>
<th>$\theta_1^L$</th>
<th>$\theta_1^U$</th>
<th>$\theta_2^L$</th>
<th>$\theta_2^U$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aramingo</td>
<td>0.714</td>
<td>0.826</td>
<td>0.042</td>
<td>0.080</td>
</tr>
<tr>
<td>Belmont</td>
<td>0.576</td>
<td>0.690</td>
<td>0.896</td>
<td>1</td>
</tr>
<tr>
<td>Bristol</td>
<td>0.690</td>
<td>0.729</td>
<td>0.278</td>
<td>0.278</td>
</tr>
<tr>
<td>Byberry</td>
<td>0.917</td>
<td>0.953</td>
<td>0.011</td>
<td>0.014</td>
</tr>
<tr>
<td>Frankford</td>
<td>1</td>
<td>1</td>
<td>0.499</td>
<td>0.499</td>
</tr>
<tr>
<td>Germantown</td>
<td>1</td>
<td>1</td>
<td>0.035</td>
<td>0.036</td>
</tr>
<tr>
<td>Horsham</td>
<td>1</td>
<td>1</td>
<td>0.263</td>
<td>0.263</td>
</tr>
<tr>
<td>Kensington</td>
<td>1</td>
<td>1</td>
<td>0.010</td>
<td>0.010</td>
</tr>
<tr>
<td>Manayunk</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Moreland</td>
<td>1</td>
<td>1</td>
<td>0.007</td>
<td>0.007</td>
</tr>
<tr>
<td>Oxford</td>
<td>0.757</td>
<td>0.805</td>
<td>0.010</td>
<td>0.012</td>
</tr>
<tr>
<td>Passyunk</td>
<td>0.791</td>
<td>0.791</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Richmond</td>
<td>1</td>
<td>1</td>
<td>0.152</td>
<td>0.152</td>
</tr>
<tr>
<td>Roxborough</td>
<td>0.834</td>
<td>0.834</td>
<td>0.007</td>
<td>0.007</td>
</tr>
<tr>
<td>Spring Garden</td>
<td>0.472</td>
<td>0.472</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

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. $\theta_1^L$ and $\theta_1^U$ are the lower bound and upper bounds for Stage 1, respectively; $\theta_2^L$ and $\theta_2^U$ are the lower and upper bounds for Stage 2, respectively.

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

<table>
<thead>
<tr>
<th>DMU</th>
<th>$\theta_1^*$</th>
<th>$\theta_2^*$</th>
<th>Ranking</th>
<th>$\alpha_1$</th>
<th>$\beta_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aramingo</td>
<td>0.773</td>
<td>0.042</td>
<td>0.032</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>Belmont</td>
<td>0.690</td>
<td>0.896</td>
<td>0.618</td>
<td>3</td>
<td>0.354</td>
</tr>
<tr>
<td>Bristol</td>
<td>0.729</td>
<td>0.278</td>
<td>0.203</td>
<td>7</td>
<td>0.354</td>
</tr>
<tr>
<td>Byberry</td>
<td>0.917</td>
<td>0.011</td>
<td>0.010</td>
<td>13</td>
<td>1</td>
</tr>
<tr>
<td>Frankford</td>
<td>1</td>
<td>0.499</td>
<td>0.499</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Germantown</td>
<td>1</td>
<td>0.036</td>
<td>0.036</td>
<td>9</td>
<td>0.432</td>
</tr>
<tr>
<td>Horsham</td>
<td>1</td>
<td>0.263</td>
<td>0.263</td>
<td>6</td>
<td>0.275</td>
</tr>
<tr>
<td>Kensington</td>
<td>1</td>
<td>0.010</td>
<td>0.010</td>
<td>12</td>
<td>0.113</td>
</tr>
<tr>
<td>Manayunk</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.113</td>
</tr>
<tr>
<td>Moreland</td>
<td>1</td>
<td>0.007</td>
<td>0.007</td>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td>Oxford</td>
<td>0.844</td>
<td>0.012</td>
<td>0.010</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>Passyunk</td>
<td>0.791</td>
<td>1</td>
<td>0.791</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Richmond</td>
<td>1</td>
<td>0.152</td>
<td>0.152</td>
<td>8</td>
<td>0.233</td>
</tr>
<tr>
<td>Roxborough</td>
<td>0.834</td>
<td>0.007</td>
<td>0.006</td>
<td>15</td>
<td>0.233</td>
</tr>
<tr>
<td>Spring Garden</td>
<td>0.472</td>
<td>1</td>
<td>0.472</td>
<td>5</td>
<td>0.233</td>
</tr>
<tr>
<td>Mean</td>
<td>0.870</td>
<td>0.348</td>
<td>0.274</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.158</td>
<td>0.415</td>
<td>0.326</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min</td>
<td>0.472</td>
<td>0.007</td>
<td>0.006</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Max</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

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. $\theta_1^*$ is the efficiency of Stage 1; $\theta_2^*$ is the efficiency of Stage 2; $\theta_*$ is the overall efficiency; $\alpha_1$ is the portion of $\theta_1^*$ consumed in Stage 1; $\beta_1$ is the portion of $\theta_2^*$ 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


