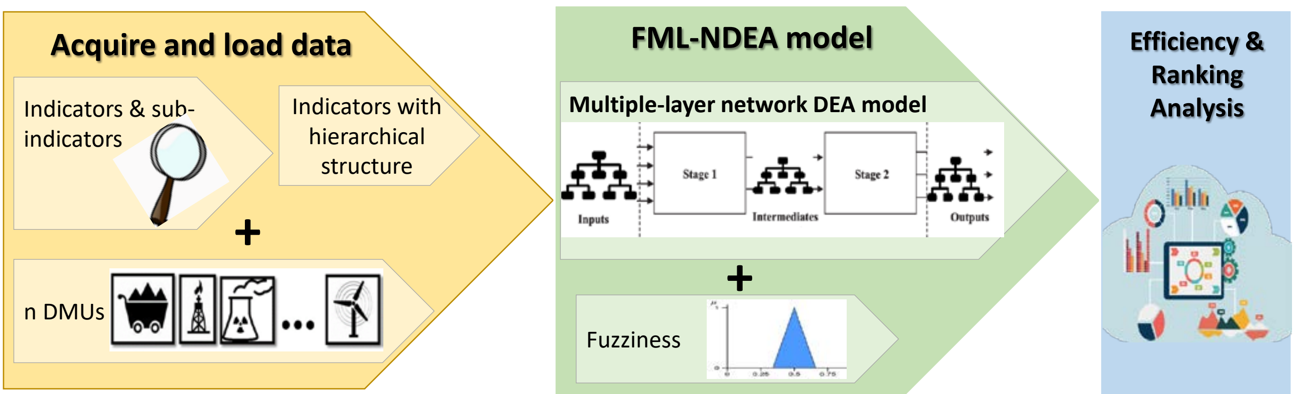
A Generalized Fuzzy Multiple-Layer NDEA: An Application to Performance-Based Budgeting

Abstract

Network data envelopment analysis (NDEA) is capable of considering operations and interdependence of a system's component processes to measure efficiencies. There are numerous performance evaluation applications in which some indicators have hierarchical structures with a considerable number of sub-indicators. This problem of ignoring the hierarchical structure of indicators weakens the discrimination power of NDEA models and may result in inaccurate efficiency scores. In this paper we propose a generalized fuzzy Multiple-Layer NDEA (GFML-NDEA) model and GFML-NDEA-based composite indicators (GFML-NDEA-CI) to incorporate the hierarchical structures of indicators in the ambit of the particular two-stage NDEA models. To demonstrate the usefulness of the GFML-NDEA-CI model proposed, its application was tested by evaluating the efficiency of the performance-based budgeting (PBB) system in 14 governmental agencies in Iran. The comparative analysis results obtained from the GFML-NDEA-CI (multi-layer) model with those from the single-layer fuzzy NDEA-CI model indicate that the number of efficient decision-making units (DMUs) in the one-layer model is eight, whereas it is solely one DMU in the multi-layer model. The discrimination power of the multi-layer model proposed is significantly increased by observing that standard deviation of efficiency scores are increased by 41%, 61%, and 84% for possibility levels 0, 0.5, and 1, respectively. This is obtained while reducing information entropy, thus suggesting that the proposed model yields more reliable scores.

Keywords: DEA; Fuzzy logic; Performance-based budgeting; Maturity model; Network structure; Hierarchical structure.

Graphical abstract



1. Introduction

Data envelopment analysis (DEA) was initially proposed by Farrell [1] to measure the efficiency of a system as a black box with a single process, ignoring its internal structure. However, many empirical studies indicate that to evaluate the performance of a system, it is important to consider the operations of its component processes. Network DEA (NDEA) models consider the entire system composed of different component processes or stages with intermediate flows among the processes and their own inputs and outputs. For studies that discuss NDEA models, one can refer to Kao [2,3]. Some works introduce new NDEA models to measure efficiencies under given conditions or present the technical properties of such models, while others apply the current models to real-world performance evaluation problems [2].

There are a large number of performance evaluation applications such as of banking systems [4–6], the insurance industry [7,8], airline industry [9,10], health care sector [11], road safety performance [12], green supply chain management [13], sustainable supply chain management [14,15], tourist hotel economic efficiency [16], and the performance-based budgeting (PBB) maturity model [17] that deal with many input and/or output indicators belonging to different categories. In addition, there are indicators that are highly correlated or have similar properties. In these applications, the high dimensionality of indicators weakens the discrimination power of the DEA models [18].

This problem of ignoring the hierarchical structure of indicators is addressed in traditional DEA models by initially introducing two-level DEA models by Meng et al. [19] and after by introducing a generalized Multilayer DEA (MLDEA) model by Shen et al. [12]. Nevertheless, the research papers mentioned above (i.e., [12,19]) are only limited to the traditional black box, single-process DEA models that do not necessarily meet the needs of today's performance evaluation applications taking into account the system's internal structure.

Practitioners often wish to consider many indicators in the network structure of their underlying system to carry out a relatively comprehensive performance evaluation in real world applications without losing the discrimination power of the model by removing less important indicators. Instead of aggregating or ignoring some indicators and their hierarchical structures in an NDEA model, it is reasonable to take a holistic approach by developing a Multiple-Layer NDEA model considering influential indicators and hierarchies with their sub-indicators.

For example, in banking system evaluations [4,6], as long as the *deposits* indicator is considered as a consolidation of four different sub-indicators (consumer checking, consumer saving, business checking, and business savings) by Wang et al. [6], it is recommended that other researchers disaggregate this indicator to identify inefficiencies better and understand how to improve banking efficiency. In performance evaluation of tourist hotels [20], considering only the number of rooms provided for rent to guests while ignoring the size and quality of the rooms may give us misleading results. In supply chain management evaluation [13,14], consolidating all types of products with different sizes and prices as one indicator (known as the number of products transported from supplier to manufacturer, from manufacturer to distributer, and from distributer to customer) is an oversimplified consideration that may result in inaccurate efficiency scores for ranking Decision-Making Units (DMUs). This is actually the case when we attempt to evaluate the PBB system of governmental agencies in Iran. In the PBB maturity model presented by Amini et al. [17], many different outputs should be measured in the evaluation process to provide relatively comprehensive budgeting system profiles of these agencies.

In this paper, we depart from previous researches and develop a generalized fuzzy Multiple-Layer NDEA model to handle vagueness nested in performance indicators within the ambit of hierarchical structures where a given indicator is a composite index of its respective sub-indicators and so on. While applied to the particular case of a two-stage NDEA within the ambit of Iranian agencies, two distinct and comprehensive fuzzy modeling approaches are presented: (i) the generalized fuzzy Multiple-Layer NDEA (GFML-NDEA) and (ii) the GFML-NDEA based Composite Indicators (GFML-NDEA-CI). While the first model addresses the traditional network issue where outputs from the previous stages serve as inputs for the subsequent ones observing the productive structure logic, the second one is focused on constructing a composite indicator (CI) by adding-up sub-indicators from previous stages observing specific weights at each layer. In other words, while the first model unveils the network productive process in an exhaustive fashion by detailing its tree structure as formed by numerous micro-activities either linked in series or parallel, the second model allows the aggregation of individual indicators computed in parallel at a given layer into a CI that represents the subsequent level while observing a series perspective. To demonstrate the applicability of the GFML-NDEA-CI model, a PBB application is performed for evaluating the efficiencies of 14 governmental agencies in Iran. Specifically within the ambit of this application of observing a PBB maturity model, the particular case of two stages is addressed, namely PBB system capabilities and PBB system results. These stages or processes are formed by the aggregation of individual indicators that observe a tree structure with branches and leaves.

The remainder of the paper is organized as follows. After a brief review of the DEA-based CI, NDEA, MLDEA, and fuzzy DEA models in Section 2, the mathematical formulation of the GFML-NDEA model and the GFML-NDEA-CI model are presented in Section 3. We apply the proposed GFML-NDEA-CI model to a PBB problem in Section 4. The experimental results from the fuzzy GFML-NDEA-CI model and its comparison with the fuzzy NDEA-based CI model with a single layer are discussed in Section 5. Managerial implications are presented in Section 6 and, finally, Section 7 concludes the paper.

2. Literature Review

2.1. Performance Measurement by the DEA-Based CI Model

Recent progress in the development of CIs includes both decision-matrix methods for supervised or unsupervised data reduction (e.g., principal component analysis, independent component analysis, factor analysis, non-linear matrix factorization, neural networks, and rough set theory) and pairwise comparison methods to derive the partial relative importance of CIs (e.g., analytical hierarchy process and fair-share ratios). A common feature among these methods is the assumption of uniform indicator weights for all DMUs under study, thus ignoring that indicator weights may vary among DMUs for the purpose of performance improvement. While uniform indicator weights help establish an unbiased DMU rank, they make it difficult to ascertain about unit-specific performance predictors.

In this respect, DEA, a data-oriented linear programming method, offers several advantages over other CI methods. First, DEA can be used to combine multiple indicators without a priori knowledge about their weight trade-offs at the DMU level. Second, DEA evaluates the relative performance of DMUs to ensure that each unit obtains the best possible set of indicator weights [21]. Any other possible set of weights would produce a lower (i.e., less favorable) composite indicator score.

Basic DEA models have traditionally been used to evaluate the relative performance of a set of DMUs based on multiple inputs and outputs capable of describing a productive process. However, DEA has also been applied more generally as a tool for multi-dimensional comparisons of DMUs. Thus, to use DEA for CI construction, i.e., combining a set of individual indicators into one overall indicator, the basic DEA optimization is constrained to have constant or equal inputs with different outputs or indicators as follows [22].

$$CI_o = \max \sum_{r=1}^{s} u_r Y_{ro}$$

$$\text{s.t.} \quad \sum_{r=1}^{s} u_r Y_{rj} \leq 1\,,\; j = 1, \dots, n \qquad (1)$$

$$u_r \geq \varepsilon,\, r = 1, \dots, s$$

The subscript, *o*, refers to the DMU whose relative efficiency is to be evaluated. $Y_{rj}(r = 1, \dots, s) \in \mathbb{R}_+^s$ is the *r*th output of the *j*th DMU, u_r is considered as non-negative weights of Y_{rj}, and ε is a small non-Archimedean number.

2.2. Network DEA

In many applications, the computational scheme for efficiency measurement must be addressed as multi-stages such as network structures [23]. Conventional DEA models generally treat the DMUs without considering their internal structure [23]. Therefore, numerous network DEA models have been presented in the literature to measure the efficiency of DMUs by taking the internal structure into account [23].

The computational structure of a closed two-stage serial NDEA model presented by Wanke & Barros [24] is depicted in Figure 1.

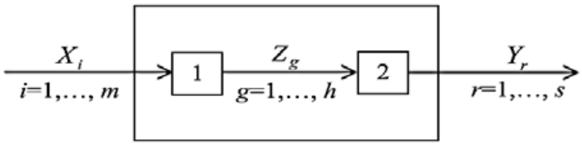


Figure 1. Closed two-stage serial NDEA model

Wanke & Barros [24] assumed a closed two-stage serial system in their NDEA model where $X_{ij}(i = 1,2,\ldots,m) \in \mathbb{R}_+^m$ and $Y_{rj}(r = 1,2,\ldots,s) \in \mathbb{R}_+^s$ stand for the first stage inputs and the second stage outputs, respectively, and $Z_{dj}(d = 1,2,\ldots,c) \in \mathbb{R}_+^c$ is intermediate measures for the j^{th} DMU. Also, v_i , w_d, u_r are considered as non-negative weights of X_{ij}, Z_{dj}, Y_{rj}, respectively. Mathematically, the efficiency score of a particular DMU_o, i.e. E_o, is obtained by solving the following basic two-stage serial NDEA model:

$$\begin{aligned}
&E_o = \max \textstyle\sum_{r=1}^{s} u_r Y_{ro} \\
&\text{s.t.} \\
&\textstyle\sum_{i=1}^{m} v_i X_{io} = 1 && i = 1,\ldots,m \\
&\textstyle\sum_{d=1}^{c} w_d Z_{dj} - \sum_{i=1}^{m} v_i X_{ij} \le 0 && j = 1,\ldots,n \\
&\textstyle\sum_{r=1}^{s} u_r Y_{rj} - \sum_{d=1}^{c} w_d Z_{dj} \le 0 && j = 1,\ldots,n \\
&u_r, v_i \ge 0
\end{aligned} \tag{2}$$

2.3. Multiple-Layer DEA (MLDEA)

The evolution of DEA models over time has pushed researchers to develop various models with hierarchical or the so-called “Multiple-Layer” structures [12]. Meng et al. [19] developed a non-linear two-layer DEA approach for the performance evaluation of research institutes and then introduced a two-layer structure for the research outputs. For the linearization of such a model, pairwise decision-making techniques are used such as AHP for weighting the internal weights within categories. Later on Kao [25] defined new variables to propose a linear two-layer DEA model. However, Meng et al.’s and Kao’s models only include two layers of indicators without providing a solution for modeling general structures with more than two layers [12]. Shen, Ruan et al.[12] introduced the generalized linear MLDEA model by studying the models proposed by Meng et al. [19] and Kao [25]. The advantage of this MLDEA model is the endogenous determination of the indicator weights [12]. The hierarchical structure of Shen et al.'s MLDEA model is depicted in Figure 2.

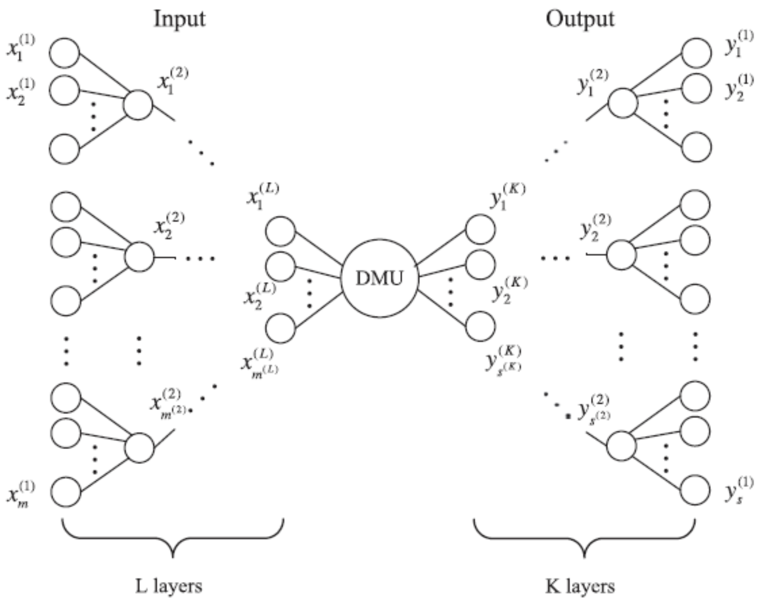


Figure 2. An MLDEA hierarchical structure [12]

The mathematical model of Shen et al.'s generalized MLDEA approach [12] is given in (3).

$$
\begin{array}{l}
Max\ E_0=\sum_{f_1=1}^{s}\hat{u}_{f_1}Y_{f_10} \\
s.t. \\
\sum_{g_1=1}^{m}\hat{v}_{g_1}X_{g_10}=1, \\
\sum_{f_1=1}^{S}\hat{u}_{f_1}Y_{f_1j}-\sum_{g_1=1}^{m}\hat{v}_{g_1}X_{g_1j}\le 0, \quad j=1,\dots,n \\
\sum_{f_1\in A_{f_k}^{(k)}}^{s^{(k)}}\hat{u}_{f_1}=u_{f_k}, \quad f_1=1,\dots,\ s, \qquad f_k=1,\dots,\ s^{(k)} \\
\sum_{g_1\in B_{g_L}^{(L)}}^{m^{(L)}}\hat{v}_{g_1}=v_{g_L}, \quad g_1=1,\dots,\ m, \qquad g_L=1,\dots,\ m^{(l)} \\
\sum_{f_1\in A_{f_k}^{(k)}}\hat{u}_{f_1}\Big/\sum_{f_1\in A_{f_{k+1}}^{(k+1)}}\hat{u}_{f_1}=p_{f_k\ f_k\in A_{f_{k+1}}^{(k+1)}}^{(k)} \qquad f_k=1,\dots,\ s^{(k)},\ k=1,\dots,K-1 \\
\sum_{g_1\in B_{g_L}^{(L)}}\hat{v}_{g_1}\Big/\sum_{g_1\in B_{g_{L+1}}^{(L+1)}}\hat{v}_{g_1}=q_{g_l\ g_l\in B_{g_{l+1}}^{(L+1)}}^{(l)} \qquad g_l=1,\dots,\ m^{(l)},\ l=1,\dots,L-1 \\
u_{f_k},v_{g_L}\ge\varepsilon,\ f_k=1,\dots,\ s^{(k)},\ g_L=1,\dots,\ m^{(l)} \\
p_{f_k\ f_k\in A_{f_{k+1}}^{(k+1)}}^{(k)}\ge\xi,\ f_k=1,\dots,\ s^{(k)},\ k=1,\dots,K-1 \\
q_{g_l\ g_l\in B_{g_{l+1}}^{(L+1)}}^{(l)}\ge\xi,\ g_l=1,\dots,\ m^{(l)},\ l=1,\dots,L-1 \\
\hat{u}_{f_1},\hat{v}_{g_1}\ge\xi^{(k-1)}\varepsilon,\ f_k=1,\dots,s^{(k)},\ g_1=1,\dots,m
\end{array}
\tag{3}
$$

Where Y_{f_kj} and X_{g_lj} are the f^{th} output and the g^{th} input, respectively, for the j^{th} DMU. A system is assumed with hierarchical structure which includes K and L layers of outputs and inputs, respectively, as shown in Figure 2. Let s^k be the number of output categories in the k^{th} layer ($k=1,2,\dots,K$), where $s^{(1)}=s$, and $m^{(l)}$ is the number of input categories in the l^{th} layer ($l=1,2,\dots,L$), where $m^{(1)}=m$ [12]. $p_{f_k}^{(k)}$ and $q_{g_l}^{(l)}$ are the internal weights related to the indicators of the f^{th} and g^{th} categories in the k^{th} output and l^{th} input layers, respectively. $u_{f_k}$is the weight given to the f^{th}

output in the K^{th} layer, $f_K = 1, \dots, s^{(K)}$, and v_{g_l} is the weight given to the g^{th} input in the L^{th} layer, $g_L = 1,\dots,m^{(L)}$. Accordingly, $u_{f_k} = \sum_{f_1 \in A_{f_k}^{(k)}}^{s} \hat{u}_{f_1}$ and $\sum_{g_1 \in B_{g_L}^{(L)}}^{m} \hat{v}_{g_1} = v_{g_L}$ are achieved by summing up the weights of the indicators in each category of each layer (i.e., $p_{f_k}^{(k)}$ and $q_{g_l}^{(l)}$) [12].

2.4. Fuzzy DEA

2.4.1. Preliminaries

In this subsection, several definitions are given in relation to the fuzzy set theory that will be applied in the rest of the paper.

Definition 1. [26] A triangular fuzzy number (TFN) $\tilde{A} = (a^l, a^m, a^u)$ is a non-negative if and only if $a^l \geq 0$. The set of all these TFNs is denoted by $TF(R)^+$.

Definition 2. Suppose $F(R)$ be the set of all fuzzy numbers, $\tilde{A}$ be a fuzzy number, and $[A_\alpha^L, A_\alpha^U]$ be the α-cut of fuzzy number $\tilde{A}$. The following linear ranking function is proposed by Pourmahmoud & Bafekr Sharak [27]:

$$F: F(R) \to R$$

$$F(\tilde{A}) = \frac{1}{2}\int_0^1 (A_\alpha^L + A_\alpha^U)\, d\alpha. \tag{4}$$

If $\tilde{A} = (a^l; a^m;\ a^u)$ is a TFN, then $F(\tilde{A}) = \frac{a^l + 2a^m + a^u}{4}$.

Definition 3. Let $\tilde{A}$ be a TFN in the form of $\tilde{A} = (a^l; a^m;\ a^u)$. Therefore it will be a symmetrical triangular fuzzy number if and only if $a^m - a^l = a^u - a^m = \alpha$. Also, $\tilde{A}$ could be defined with center $A = a^m$ and spread α.

2.4.2. Dealing with uncertain data

Addressing uncertain data in DEA has been a challenge for many researchers. In standard DEA models it is assumed that the indicator data collected are certain. If some indicator data are uncertain or vague, then it is no longer possible to use standard DEA models. By investigating the theoretical background of DEA models, two approaches are found to handle uncertain indicator data: imprecise DEA and fuzzy DEA [28]. Guo & Tanaka [29] and Guo [30] defined a fuzzy DEA model by considering both the input and output as follows:

$$Max \quad \lambda_1 \sum_{r=1}^{s} u_r(Y_{ro} - (1-pl)\alpha_{ro}) + \lambda_2 \sum_{r=1}^{s} u_r(Y_{ro} + (1-pl)\alpha_{ro}) \tag{5}$$

$$\begin{aligned}
& s.t. \\
& \sum_{i=1}^{m} v_i X_{io} \geq g_o && i = 1, \ldots, m \\
& \sum_{r=1}^{s} u_r(Y_{rj} - (1-pl)\alpha_{rj}) \leq \sum_{i=1}^{m} v_i(X_{ij} - (1-pl)b_{ij}) && j = 1, \ldots, n \\
& \sum_{r=1}^{s} u_r(Y_{rj} + (1-pl)\alpha_{rj}) \leq \sum_{i=1}^{m} v_i(X_{ij} + (1-pl)b_{ij}) && j = 1, \ldots, n \\
& u_r, v_i \geq \varepsilon && i = 1, \ldots, m; r = 1, \ldots, s
\end{aligned}$$

Therefore, λ_1 and λ_2 are defined for providing pessimistic, indifferent, and optimistic situations with $\lambda_1 \geq 0$, $\lambda_2 \geq 0$ and $\lambda_1 + \lambda_2 = 1$. Three situations are usually considered, which are pessimistic if $\lambda_1 = 1$, optimistic if $\lambda_2 = 1$, and indifferent if $\lambda_1 = \lambda_2 = 0.5$.

Since $max \left[\frac{b_{01}}{X_{01}}, \ldots, \frac{b_{0s}}{X_{0s}}\right] \leq e$, then $\frac{b_{0k}}{X_{0k}} \leq e$ holds. Considering n DMUs, e is taken as $e = Max_{j=1,\ldots,n} \left(Max_{k=1,\ldots,s}\, b_{jk} \,/\, X_{jk}\right)$. Then, by substituting e in the following model, $\mathrm{g_o}$ is calculated for each DMU as follows [30]:

$$g_o = max \quad \sum_{i=1}^{m} v_i b_{io} \tag{6}$$

$$\begin{aligned}
& s.t. \\
& \sum_{i=1}^{m} v_i(X_{io} - (1-pl)b_{io}) = 1 - (1-pl) * e \\
& \sum_{i=1}^{m} v_i(X_{io} + (1-pl)b_{io}) \leq 1 + (1-pl) * e \\
& v_i \geq \varepsilon && i = 1, \ldots, m
\end{aligned}$$

In Model (5), *pl* is a parameter in [0,1] which indicates the uncertainty level of the decision maker where the smaller *pl* means more uncertainty. For more detail refer to [30].

3. A Fuzzy Multiple-Layer Two-Stage Serial NDEA-based CI

In this section, first the particular two-stage NDEA model was developed, which is applicable for measuring the efficiency of any two-stage serial system in which some indicators present hierarchical structures with Multilayers. Fuzzy logic is then introduced into the modeling to deal with vagueness in variable measurements such as those derived from managerial perceptions or preferences. Subsequently, the GFML-NDEA applied for the particular two-stage structure is presented for the case where it is necessary to aggregate CIs. This was done by imposing a specific set of weight constraints as shown in Section 3.4.

3.1. A Multiple-Layer Two-Stage Serial NDEA Model

Given the two-stage serial NDEA model proposed by Wanke & Barros [24] and taking advantage of the DEA model logic with the Multiple-Layer structure proposed by Shen et al. [12], the generalized Multiple-Layer NDEA (GML-NDEA) model can be developed as Figure 3.

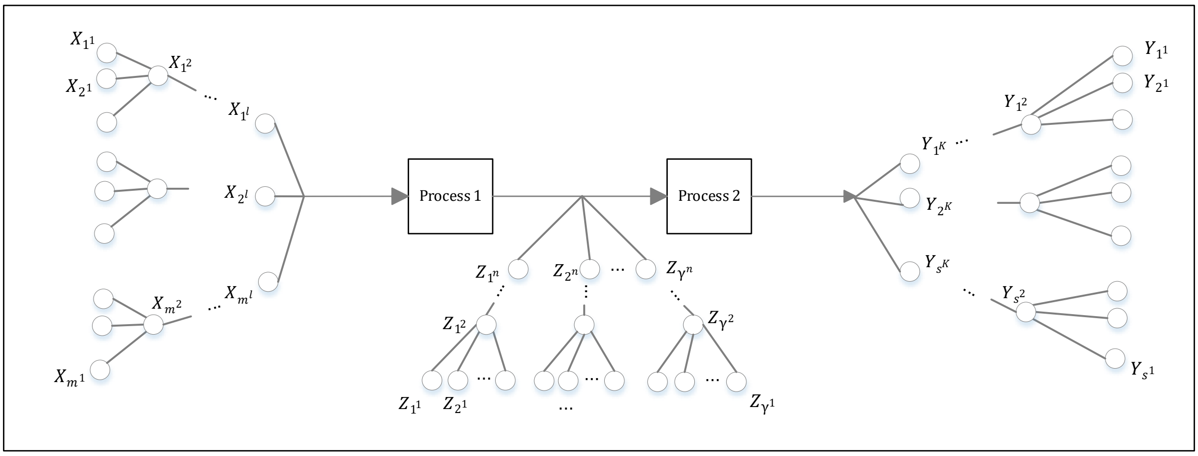


Figure 3. A hierarchical structure of the GML-NDEA model

In a system with a hierarchical structure with *G* layers of inputs, *Γ* layers of intermediates, and *F* layers of outputs as depicted in Figure 3, in the GML-NDEA model, $X_{g_l j}$,$Z_{\gamma_n j}$, and $Y_{f_k j}$ are the g^{th}, γ^{th}, and f^{th} input, intermediate and output indicators in DMU$_j$ on the l^{th}, n^{th}, and k^{th} layers of hierarchy, respectively. Furthermore, $q_{g_l}^{(l)}$, $o_{\gamma_n}^{(n)}$, and $p_{f_k}^{(k)}$ are the internal weights related to the indicators of the g^{th}, γ^{th}, and f^{th} categories on the l^{th}, n^{th}, and k^{th} layers of hierarchy for the input, intermediate, and output indicators, respectively. Where u_{f_k} is the weight given to the f^{th} output in the k^{th} layer (i.e., the final layer), $f_K = 1, \ldots, s^{(K)}$, v_{g_l} is the weight given to the g^{th} input in the l^{th} layer, $g_L = 1, \ldots, m^{(L)}$, and w_{γ_n} is the weight given to the γ^{th} intermediate indicator in the n^{th} layer and $\gamma_N = 1, \ldots, \beta^{(n)}$.

By summing up the weights of the indicators in each category of each layer (i.e., $p_{f_k}^{(k)}$, $o_{\gamma_n}^{(n)}$, and $q_{g_l}^{(l)}$), we get $u_{f_k} = \sum_{f_1 \in A_{f_k}^{(k)}}^{s} \hat{u}_{f_1}$, $w_{\gamma_n} = \sum_{\gamma_1 \in C_{\gamma_n}^{(n)}}^{\beta} \widehat{w}_{\gamma_1}$ and $v_{g_L} = \sum_{g_1 \in B_{g_L}^{(L)}}^{m} \hat{v}_{g_1}$. Accordingly, Model (7) depicts the GML-NDEA model.

$$E_o = Max \sum_{f_1=1}^{S} \hat{u}_{f_1} Y_{f_1 o} \tag{7}$$

$s.t.$

$$\sum_{g_1=1}^{m} \hat{v}_{g_1} X_{g_1 o} = 1 \tag{7i}$$

$$\sum_{f_1=1}^{s} \hat{u}_{f_1} Y_{f_1 j} - \sum_{g_1=1}^{m} \hat{v}_{g_1} X_{g_1 j} \le 0 \qquad j = 1, \dots, n \tag{7ii}$$

$$\sum_{f_1=1}^{s} \hat{u}_{f_1} Y_{f_1 j} - \sum_{\gamma_1=1}^{\beta} \hat{w}_{\gamma_1} Z_{\gamma_1 j} \le 0 \qquad j = 1, \dots, n \tag{7iii}$$

$$\sum_{\gamma_1=1}^{\beta} \hat{w}_{\gamma_1} Z_{\gamma_1 j} - \sum_{g_1=1}^{m} \hat{v}_{g_1} X_{g_1 j} \le 0 \qquad j = 1, \dots, n \tag{7iv}$$

$$p^{(k)}_{f_k \; f_k \in A^{(k+1)}_{f_{k+1}}} = \frac{\sum_{f_1 \in A^{(k)}_{f_k}}^{S} \hat{u}_{f_1}}{\sum_{f_1 \in A^{(k+1)}_{f_{k+1}}}^{S} \hat{u}_{f_1}} \qquad f_k = 1, \dots, \; s^{(k)}, \; k = 1, \dots, K-1 \tag{7v}$$

$$q^{(L)}_{g_L \; g_L \in B^{(L+1)}_{g_{L+1}}} = \frac{\sum_{g_1 \in B^{(L)}_{g_L}}^{m} \hat{v}_{g_1}}{\sum_{g_1 \in B^{(L+1)}_{g_{L+1}}}^{m} \hat{v}_{g_1}} \qquad g_l = 1, \dots, \; m^{(l)}, \quad l = 1, \dots, L-1 \tag{7vi}$$

$$o^{(n)}_{\gamma_n \; \gamma_n \in C^{(n+1)}_{\gamma_{n+1}}} = \frac{\sum_{\gamma_1 \in C^{(n)}_{\gamma_n}}^{\beta} \hat{w}_{\gamma_1}}{\sum_{\gamma_1 \in C^{(n+1)}_{\gamma_{n+1}}}^{\beta} \hat{w}_{\gamma_1}} \qquad \gamma_n = 1, \dots, \beta^{(n)} \quad n = 1, \dots, N-1 \tag{7vii}$$

$$\sum_{f_1 \in A^{(k)}_{f_k}}^{s} \hat{u}_{f_1} = u_{f_k} \qquad f_k = 1, \dots, \; s^{(k)}, \quad k = 1, \dots, K-1 \tag{7viii}$$

$$\sum_{g_1 \in B^{(L)}_{g_L}}^{m} \hat{v}_{g_1} = v_{g_L} \qquad g_l = 1, \dots, \; m^{(l)}, \quad l = 1, \dots, L-1 \tag{7ix}$$

$$\sum_{\gamma_1 \in C^{(n)}_{\gamma_n}}^{\beta} \hat{w}_{\gamma_1} = w_{\gamma_n} \qquad \gamma_n = 1, \dots, \beta^{(n)} \quad n = 1, \dots, N-1 \tag{7x}$$

$$u_{f_k}, v_{g_l}, w_{\gamma_n} \ge \varepsilon \tag{7xi}$$

$$p^{(k)}_{f_k \; f_k \in A^{(k+1)}_{f_{k+1}}} \ge \xi \tag{7xii}$$

$$q^{(l)}_{g_l \; g_l \in B^{(L+1)}_{g_{l+1}}} \ge \xi \tag{7xiii}$$

$$o^{(n)}_{\gamma_n \; \gamma_n \in C^{(n+1)}_{\gamma_{n+1}}} \ge \xi \tag{7xiv}$$

$$\hat{u}_{f_1}, \hat{v}_{g_1}, \hat{w}_{\gamma_1} \ge \xi^{(k-1)} \varepsilon \tag{7xv}$$

Where constraint (7ii) corresponds to the system as redundant and can be omitted. Constraints (7iii) and (7iv) correspond to first and second processes, respectively. Constraints $(7\text{v} - 7\text{vii})$ show the layer weights of each indicator category, and constraints $(7\text{viii} - 7\text{x})$ represent the sum of indicator weights of each category associated with the corresponding layer. The ε and ξ in constraints (7xi-7xv) are small non-Archimedean numbers that are imposed to model for preventing the DMU to assign a weight of zero to unfavorable indicators.

Where E_o is the overall efficiency level of two-stage process for DMU_o. Assuming that Model (7) yields a unique solution, the efficiencies for the first and second stages are $E_o^1 = \sum_{\gamma_1=1}^{\beta} \hat{w}_{\gamma_1} Z_{\gamma_1 o}$ and

$E_o^2 = \sum_{f_1=1}^{s} \hat{u}_{f_1} Y_{f_1 o} / \sum_{\gamma_1=1}^{\beta} \hat{w}_{\gamma_1} Z_{\gamma_1 o}$, respectively. The overall efficiency level could be calculated by product of the individual efficiency levels for each stage, i.e. $E_o = E_o^1 \times E_o^2$.

3.2. A Fuzzy GML-NDEA Model

The input and output indicators in Model (7) are assumed to be a set of quantitative data for the performance assessment process. However, in many applications some indicators are uncertain or vague, and in this case the fuzzy set theory, as briefly explained in Section 2.4, has been found as a valuable approach to tackle imprecision and vagueness in a DEA framework. The approach proposed by Shen et al. [34] is employed in this paper in order to deal with the uncertain data. Model (8) is similar to Model 7 with the exception that in Model 8 each indicator is presented as a fuzzy number. In developing a fuzzy GML-NDEA (GFML-NDEA) model, the qualitative indicator data are represented as fuzzy numerical values by simply considering the crisp indicators $X_{g_l j}$,$Z_{\Upsilon_n j}$, and $Y_{f_k j}$ as fuzzy indicators $\tilde{X}_{g_l j}$,$\tilde{Z}_{\Upsilon_n j}$, and $\tilde{Y}_{f_k j}$, respectively. Where they are assumed to be symmetrical triangular fuzzy numbers as represented by pairs with corresponding centers and spreads, $\tilde{X}_{g_l j} = (X_{g_1 j}, b_{g_1 j})$, $\tilde{Z}_{\Upsilon_n j} = (Z_{\gamma_1 j}, d_{\gamma_1 j})$ and $\tilde{Y}_{f_k j} = (Y_{f_1 j}, \alpha_{f_1 j})$ where $X_{g_1 j}$, $Z_{\gamma_1 j}$ and $Y_{f_1 j}$ are the normalized value for related indicators and $b_{g_1 j}$, $d_{\gamma_1 j}$ and $\alpha_{f_1 j}$ are their spread, respectively. The rest of the parameters and variables are similar to Model 7. The resulted GFML-NDEA model has the form of a fuzzy linear programming problem with fuzzy coefficients both in the objective function and in the constraints. Considering Model (7), the GFML-NDEA model is transformed to the following crisp linear programming problem.

$$E_o = max\ \lambda_1 \sum_{f_1=1}^{S} \hat{u}_{f_1}(Y_{f_1 o} - (1-pl)\alpha_{f_1 o}) + \lambda_2 \sum_{f_1=1}^{S} \hat{u}_{f_1}(Y_{f_1 o} + (1-pl)\alpha_{f_1 o}) \qquad (8)$$

$$s.t.$$

$$\sum_{g_1=1}^{m} \hat{v}_{g_1} X_{g_1 o} \geq g_o$$

$$\sum_{f_1=1}^{S} \hat{u}_{f_1}(Y_{f_1 j} + (1-pl)\alpha_{f_1 j}) - \sum_{g_1=1}^{m} \hat{v}_{g_1}(X_{g_1 j} + (1-pl)b_{g_1 j}) \leq 0 \qquad j = 1, \dots, n$$

$$\sum_{\gamma_1=1}^{\beta} \hat{w}_{\gamma_1}(Z_{\gamma_1 j} + (1-pl)d_{\gamma_1 j}) - \sum_{g_1=1}^{m} \hat{v}_{g_1}(X_{g_1 j} + (1-pl)b_{g_1 j}) \leq 0 \qquad j = 1, \dots, n$$

$$\sum_{f_1=1}^{S} \hat{u}_{f_1}(Y_{f_1 j} + (1-pl)\alpha_{f_1 j}) - \sum_{\gamma_1=1}^{\beta} \hat{w}_{\gamma_1}(Z_{\gamma_1 j} + (1-pl)d_{\gamma_1 j}) \leq 0 \qquad j = 1, \dots, n$$

$Cnstraints$ 7v to 7xii

Here, pl is the possibility level that is determined by the decision-makers, so naturally $pl \in [0,1]$ is expected. In practice, the given possibility level by decision-makers represents their attitude toward uncertainty. When $pl = 1$, the fuzzy data are supposed to be treated as crisp and the same indicator scores are obtained for each DMU regardless of whether the decision-makers take a pessimistic, indifferent, or optimistic consideration. When the given value to pl becomes lower than 1, the decision-makers are supposed to be more cautious with respect to data ambiguity and vagueness. As a consequence, a wider range of indicator scores can be derived. In such a way, the uncertainties or vagueness related to the indicators measured are captured by the model. The fuzzy indicator scores are achieved in accordance with the various possibility levels (pl) using the symmetric triangular fuzzy output vector $Y_o = (y_o, d_o)_L$. λ_1 and λ_2 are defined for providing pessimistic, indifferent, and optimistic situations with $\lambda_1 \geq 0$, $\lambda_2 \geq 0$ and $\lambda_1 + \lambda_2 = 1$. Three situations are usually considered, which are pessimistic if $\lambda_1 = 1$, optimistic if $\lambda_2 = 1$, and indifferent if $\lambda_1 = \lambda_2 = 0.5$. The fuzzy efficiency score of DMU_o can then be defined as $\{\sum_{f_1=1}^{S} \hat{u}^*{}_{f_1}(Y_{f_1o} - (1-pl)\alpha_{f_1o},\ \sum_{f_1=1}^{S} \hat{u}^*{}_{f_1} Y_{f_1o},\ \sum_{f_1=1}^{S} \hat{u}^*{}_{f_1}(Y_{f_1o} + (1-pl)\alpha_{f_1o}\}$, which again represents pessimistic, indifferent, and optimistic scenarios.

To calculate the value of g_o in the above serial network model, as in the case of [30], first e is taken as $e = Max_{j=1,\dots,n}\ (Max\ b_{g_1j}\ /\ X_{g_1j})$. Then, by substituting e in the following model, g_o is calculated for each DMU as follows:

$$
\begin{aligned}
& g_o = max \quad \sum_{g_1=1}^{m} \hat{v}_{g_1} b_{g_1o} \qquad (9)\\
& \text{s.t.}\\
& \sum_{g_1=1}^{m} \hat{v}_{g_1}(X_{g_1o} - (1-pl)b_{g_1o}) = 1 - (1-pl) * e\\
& \sum_{g_1=1}^{m} \hat{v}_{g_1}(X_{g_1o} + (1-pl)b_{g_1o}) \leq 1 + (1-pl) * e\\
& \hat{v}_{g_1} \geq \varepsilon \qquad\qquad g_1 = 1, \dots, m
\end{aligned}
$$

The optimization problem (9) is used to seek the maximum Z = $\hat{v}_{g_1} b_{g_1o}$ constrained by $\hat{v}_{g_1} b_{g_1o} \lesssim \tilde{1}$ with the same left endpoint as fuzzy number $\tilde{1}$ in pl-level sets. This approach can be regarded as a generalization of the procedure in which seeking a value x such that x = 1 is equivalent to finding out the biggest x subject to $x \leq 1$. For more details refer to [29].

3.3. A GFML-NDEA-based CI Model

Model (8) represents a general situation in which all processes have both inputs and outputs when there is a situation where you are looking to construct an aggregated index, or in other words, there is no input, or there is a constant or equal input (i.e. $\sum_{g_1=1}^{m} \hat{v}_{g_1} X_{g_1o}$ is equivalent with 1 and the spread of X_{g_1j}, i.e. b_{g_1j}, is equal to zero). Taking the Shen's DEA-CI model [22] into consideration, for all $pl \in [0,1]$ we have $\sum_{g_1=1}^{m} (\hat{v}_{g_1} X_{g_1j} + (1-pl) b_{g_1j}) = 1$. Accordingly, the GFML-NDEA model can be converted to a GFML-NDEA-based CI (GFML-NDEA-CI) model as follows:

$$CI_o = max\ \lambda_1 \sum_{f_1=1}^{S} \hat{u}_{f_1}(Y_{f_1 o} - (1-pl)\alpha_{f_1 o}) + \lambda_2 \sum_{f_1=1}^{S} \hat{u}_{f_1}(Y_{f_1 o} + (1-pl)\alpha_{f_1 o}) \quad (10)$$
$$s.t.$$
$$\sum_{f_1=1}^{S} \hat{u}_{f_1}(Y_{f_1 j} + (1-pl)a_{f_1 j}) \leq 1$$
$$\sum_{\gamma_1=1}^{\beta} \widehat{w}_{\gamma_1}(Z_{\gamma_1 j} + (1-pl)d_{\gamma_1 j}) \leq 1$$
$$\sum_{f_1=1}^{S} \hat{u}_{f_1}(Y_{uf_1 j} + (1-pl)a_{f_1 j}) - \sum_{\gamma_1=1}^{\beta} \widehat{w}_{\gamma_1}(Z_{\gamma_1 j} + (1-pl)d_{\gamma_1 j}) \leq 0$$
$$Constraints\ \text{7v, 7vii, 7viii, 7x, 7xi, 7xii, 7xiv, and 7xv}$$

Where the fuzzy index score of DMU_o, i.e. CI_o, it can then be defined as $\{\sum_{f_1=1}^{S} \hat{u}^*_{f_1}(Y_{f_1 o} - (1-pl)\alpha_{f_1 o}, \sum_{f_1=1}^{S} \hat{u}^*_{f_1} Y_{f_1 o}, \sum_{f_1=1}^{S} \hat{u}^*_{f_1}(Y_{f_1 o} + (1-pl)\alpha_{f_1 o}\}$, which again represents pessimistic, indifferent, and optimistic scenarios.

Definition 4. The relative efficiency score for each DMU_o is a triangular fuzzy number in a possibility level pl that can be defined in three pessimistic, indifferent, and optimistic modes as follows:

$$Fuzzy\ CI_o^{pl} = [\text{Pes}, \text{Ind}, \text{Opt}] = \left[\sum_{f_1=1}^{S} \hat{u}^*_{f_1}\big(Y_{f_1 o} - (1-pl)a_{f_1 o}\big), \sum_{f_1=1}^{S} \hat{u}^*_{f_1} Y_{f_1 o}, \sum_{f_1=1}^{S} \hat{u}^*_{f_1}\big(Y_{f_1 o} + (1-pl)a_{f_1 o}\big)\right] \quad (11)$$

Definition 5. If $Fuzzy\ CI_o^{pl}$ be the triangular fuzzy score of DMU_o in possibility level pl, based on Definition 3, $Ag\ Fuzzy\ CI_o^{pl}$ can be defined as an aggregated fuzzy score for DMU_o in possibility level pl as follows:

$$Ag\ Fuzzy\ CI_o^{pl} = \frac{\sum_{f_1=1}^{S} \hat{u}_{f_1}\big(Y_{f_1 o} - (1-pl)a_{f_1 o}\big) + 2\ \sum_{f_1=1}^{S} \hat{u}_{f_1} Y_{f_1 o} + \sum_{f_1=1}^{S} \hat{u}_{f_1}\big(Y_{f_1 o} + (1-pl)a_{f_1 o}\big)}{4} \quad (12)$$

3.4. Imposing some restrictions on weights

DEA research has suggested a wide variety of weight restriction methods (e.g., [21,31]). The four commonly applied techniques of weight restrictions include "absolute weight restrictions", "relative weight restrictions", and "ordinal weight restrictions" [12].

In this research, we added absolute weight restrictions, i.e., $L_{f_k}^{(k)} \leq p_{f_k}^{(k)} \leq U_{f_k}^{(k)}$, where $f_k \in A_{f_{k+1}}^{(k+1)}$, $k = 1, \ldots, K+1$, L and U represent the lower and upper bounds of the internal weights, respectively [12], and further incorporated them into the model.

4. Application in PBB

There are many reasons why governments want to measure performance and allocate budget based on performance. The most important reason is probably the decision to allocate scarce state resources based on which government programs yield better results and therefore deserve increases in the budget. However, the public sector is not inherently intended to achieve profitability, so without the profitability driver it is challenging to identify which programs produce benefits and which do not [32]. Performance measurements would address this problem by providing quantitative evidence demonstrating that programs serve their purposes and to what extent. Performance-based budgeting uses the performance measurement output in the budgeting process, preferably reflected in budget allocations that more precisely represent the program's relative value [32].

As a result of pressure from the central government, the PBB is often used to boost performance [33]. The Plan and Budget Organization (PBO) of Iran has also established a PBB program and in accordance with the annex to the 2013 budget deficit, it advised all entities of this program. Although organizations are showing the expected motivation to move towards the PBB model, they are progressing slowly and behind the planned schedule. It is necessary to identify, assess, and recognize the organizations that have begun using the PBB program to promote the change while also identifying good practices and share successful experiences with other organizations. To this purpose, the progress of implementing PBB in different organizations should be monitored and evaluated.

4.1. PBB Maturity Model Hierarchy

Maturity models can be used to evaluate the ability of a system in performance measurement. The PBB maturity model proposed by Amini et al. [17] is an evaluation model that seeks to assess the progress of organizations in establishing and implementing a successful budgeting system and consists of “capabilities” and “results” processes while a set of indicators is defined for each process as shown in Figure 4.

The first process assesses the capabilities of PBB sub-systems. It evaluates six main sub-systems of the budgeting system including "planning, information management, process management & documentation, costing system, performance management, and control & monitoring". The second process assesses the results of a successful deployment of a PBB system. In addition to evaluating all capabilities of a budgeting system, the extent to which the goals of that system are achieved is also considered. When defining the results indicators, the focus must be on the expectations of benefits that each organization has in the event of a full and correct implementation of the PBB systems. Accordingly, by reviewing the literature and interviewing the budgeting experts, four main result indicators were defined: "transparency & accountability”, “budget discipline”, “applying performance information”, and “applying costing information”[17].

PBB maturity measurement in Iran is usually conducted in two phases. In order to evaluate the organization's budget maturity, a PBB rating was designed by the advisory committee of the International Conference on PBB held in 2018 in Tehran, Iran. Initially, all organizations, agencies, and institutions that claimed to deploy and implement a PBB system in their organization were invited to participate in the process of evaluating their budget maturity. Accordingly, the process of assessing the maturity of the budgeting system began with announcing the readiness of 20 organizations.

In the first phase, self-declaration forms were filled out by the organizations regarding the activities carried out in the field of the budget. The forms filled out were evaluated by several trained auditors, including a few representatives from the PBO of Iran, the Supreme Audit Court (SAC) of Iran, and some budget experts in Iran. Based on the initial evaluation, only 14 organizations from different sectors, including electric energy, tourism, banking, economic planning, and social security, were able to attain the minimum standards required to enter the final evaluation stage of the budget maturity rating.

In the second phase, evaluation forms containing both qualitative and quantitative indicators were used in two parts. The first part was devoted to assessing the capabilities of a mature system complemented by a Likert-based questionnaire (six main criteria and twelve sub-criteria for each criterion) through a review of the documentation and systems as well as interviews. The second part was related to evaluating the results of a mature PBB system. Its quantitative indicators were collected through the SAC of Iran and quality indicators were filled in using the form of a questionnaire through observations and interviews.

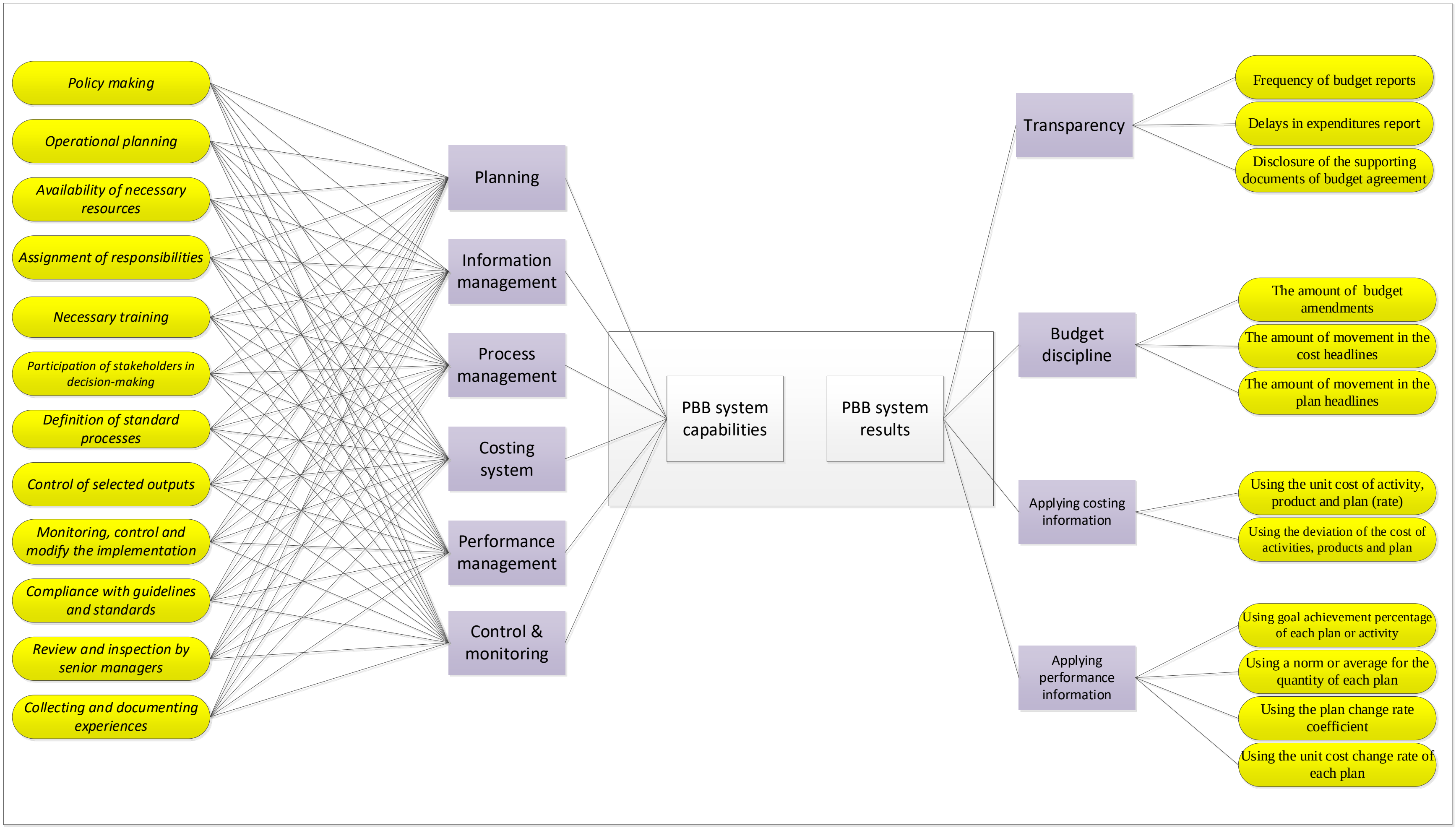


Figure 4. The hierarchal structure of the PBB maturity model proposed by Amini et al. [17]

4.2. Available Data

The data collected described above was prepared by organizing the qualitative data based on the logic introduced by [22] as a triangular fuzzy number. Accordingly, the center and spread of each piece of qualitative data $\tilde{Y}_{rj} = (Y_{rj}, \alpha_{rj})$ are listed in Table 1. Descriptive statistics and raw data for all sub-indicators are provided in Appendix A and Supplementary material, respectively.

Table 1. The symmetrical triangular fuzzy numbers for the Likert-based qualitative data

Qualitative data	1	2	3	4	5
Fuzzy number (Y_{rj}, α_{rj})	$\left(\frac{1}{5}, \frac{1}{5}\right)$	$\left(\frac{2}{5}, \frac{1}{5}\right)$	$\left(\frac{3}{5}, \frac{1}{5}\right)$	$\left(\frac{4}{5}, \frac{1}{5}\right)$	$\left(\frac{5}{5}, \frac{1}{5}\right)$

The raw data should be normalized in order to eliminate the scale differences among the sub-indicators. A considerable number of normalization methods have been introduced in the literature including rescaling, standardization, and ranking [34]. Therefore, in our study, the distance to a reference approach [35] is applied.

4.3. Weight Restrictions

The following formulas are defined to determine the lower and upper bounds of the interval weights of sub-indicators into their indicator:

$$Lower\ bound\ for\ \text{interval weights}\ of\ sub-indicator = \frac{1}{\text{Number of sub indicators in } f^{th} \text{ indicator}} \times 0.5$$

$$Upper\ bound\ for\ \text{interval weights}\ of\ sub-indicator = \frac{1}{\text{Number of sub indicators in } f^{th} \text{ indicator}} \times 2$$

As for the above restrictions, all 12 sub-indicators in each capability indicator should be in the range of [0.041667, 0.16667]. Likewise, the interval weights of all the three sub-indicators in the first and second indicators should be in the range of [0.16667, 0.6667], and the share of all sub-indicators in the third and fourth indicators should be in the ranges of [0.25, 1] and [0.125, 0.5], respectively. Also, $\varepsilon = 0.01$, $\xi = 0.1$, and $\xi \times \varepsilon = 0.001$ are imposed to avoid assigning zero weights.

5. Experimental Results

Composite efficiency scores of the PBB maturity indicator for 14 governmental agencies were computed by applying the GFML-NDEA-CI model (10), and the results are presented in Table 2. As seen, the model is solved for each possibility level $pl \in \{0, 0.5, 1\}$. Based on the defuzzification of the model, triangular fuzzy efficiency scores are calculated for each level of pl in Table 2.

Table 2. Composite efficiency indicator scores of 14 governmental agencies based on the GFML-NDEA-CI model

DMUs	Sector	Composite Indicator (CI)		
		$pl = 0$	$pl = 0.5$	$pl = 1$

DMU_{01}	Banking	{0.483, 0.605, 0.806}	{0.574, 0.651, 0.787}	{0.763, 0.763, 0.763}
DMU_{02}	Economic planning	{0.515, 0.605, 0.817}	{0.562, 0.679, 0.799}	{0.779, 0.779, 0.779}
DMU_{03}	Social security	{0.405, 0.592, 0.784}	{0.549, 0.655, 0.762}	{0.733, 0.733, 0.733}
DMU_{04}	Electric energy	{0.256, 0.407, 0.688}	{0.345, 0.476, 0.647}	{0.589, 0.589, 0.589}
DMU_{05}	Economic planning	{0.652, 0.823, 1}	{0.806, 0.902, 1}	{1, 1, 1}
DMU_{06}	Economic planning	{0.514, 0.613, 0.824}	{0.571, 0.688, 0.807}	{0.786, 0.786, 0.786}
DMU_{07}	Banking	{0.508, 0.630, 0.821}	{0.592, 0.694, 0.804}	{0.785, 0.785, 0.785}
DMU_{08}	Tourism	{0.509, 0.679, 0.816}	{0.649, 0.727, 0.793}	{0.766, 0.766, 0.766}
DMU_{09}	Tourism	{0.273, 0.538, 0.808}	{0.473, 0.631, 0.789}	{0.765, 0.765, 0.765}
DMU_{10}	Banking	{0.640, 0.820, 0.997}	{0.801, 0.898, 0.994}	{0.987, 0.987, 0.987}
DMU_{11}	Social security	{0.434, 0.597, 0.799}	{0.557, 0.653, 0.778}	{0.753, 0.753, 0.753}
DMU_{12}	Social security	{0.523, 0.712, 0.902}	{0.683, 0.788, 0.893}	{0.881, 0.881, 0.881}
DMU_{13}	Banking	{0.539, 0.665, 0.836}	{0.633, 0.722, 0.817}	{0.795, 0.795, 0.795}
DMU_{14}	Electric energy	{0.305, 0.443, 0.712}	{0.387, 0.510, 0.676}	{0.625, 0.625, 0.625}

The model results can be reported and analyzed for different levels of pl and for three maturity scores, namely the pessimistic, indifferent, and optimistic situations. As seen in Table 2, DMU_{05} is considered the top unit followed by $DMU_{10}, DMU_{12}, DMU_{13}$, respectively. As the value of reliability pl rises from 0 to 1, the fuzzy score interval (distance between the pessimistic and optimistic scores) decreases. Looking at the fifth column, for $pl = 1$, which represents complete reliability, the interval approaches zero.

5.1. Aggregating the Triple Fuzzy Scores of PBB Maturity Indicator

With regard to Definition 2, three pessimistic, indifferent, and optimistic fuzzy numbers are calculated as the final scores of the DMUs. As shown in Table 2, there are lots of PBB maturity indicator scores related to different possibility levels (pl) and three fuzzy scores for pessimistic, indifferent, and optimistic situations. To better analyze the PBB maturity indicator, according to Definition 5, the aggregated fuzzy scores of the DMUs are calculated as presented in Table 3. The aggregated fuzzy efficiency scores of overall PBB maturity indicator (Composite Indicator or CI) as well as its two processes, i.e. capabilities and results as illustrated in Section 4.1 and Figure 4, for each possibility level $pl \in \{0, 0.5, 1\}$ are depicted in Table 3.

Table 3. The aggregated fuzzy efficiency scores of overall PBB maturity indicator and its two processes, capabilities and results.

DMUs	Sector	Overall CI			Capabilities' CI			Results' CI		
		$pl = 0$	$pl = 0.5$	$pl = 1$	$pl = 0$	$pl = 0.5$	$pl = 1$	$pl = 0$	$pl = 0.5$	$pl = 1$
$\mathbf{DMU_{01}}$	Banking	0.625	0.666	0.764	0.766	0.841	0.989	0.819	0.792	0.772
$\mathbf{DMU_{02}}$	Economic planning	0.636	0.68	0.78	0.671	0.737	0.875	0.95	0.924	0.892

DMUs	Sector	Overall CI			Capabilities' CI			Results' CI		
		$pl = 0$	$pl = 0.5$	$pl = 1$	$pl = 0$	$pl = 0.5$	$pl = 1$	$pl = 0$	$pl = 0.5$	$pl = 1$
$\mathbf{DMU_{03}}$	Social security	0.594	0.656	0.734	0.594	0.656	0.734	1	1	1
$\mathbf{DMU_{04}}$	Electric energy	0.44	0.486	0.59	0.806	0.893	1	0.53	0.538	0.59
$\mathbf{DMU_{05}}$	Economic planning	0.825	0.903	1	0.825	0.903	1	1	1	1
$\mathbf{DMU_{06}}$	Economic planning	0.642	0.689	0.787	0.649	0.698	0.787	0.987	0.985	1
$\mathbf{DMU_{07}}$	Banking	0.647	0.696	0.785	0.806	0.893	1	0.804	0.779	0.785
$\mathbf{DMU_{08}}$	Tourism	0.672	0.725	0.766	0.677	0.73	0.788	0.993	0.994	0.972
$\mathbf{DMU_{09}}$	Tourism	0.54	0.631	0.765	0.62	0.683	0.765	0.843	0.918	1
$\mathbf{DMU_{10}}$	Banking	0.82	0.898	0.988	0.82	0.898	0.988	1	1	1
$\mathbf{DMU_{11}}$	Social security	0.607	0.661	0.753	0.607	0.661	0.753	1	1	1
$\mathbf{DMU_{12}}$	Social security	0.713	0.788	0.882	0.713	0.788	0.882	1	1	1
$\mathbf{DMU_{13}}$	Banking	0.677	0.724	0.796	0.677	0.724	0.796	1	1	1
$\mathbf{DMU_{14}}$	Electric energy	0.476	0.521	0.626	0.616	0.688	0.784	0.762	0.75	0.798

The aggregated score of overall CI indicates that DMU_{05} has the highest maturity. Then DMU_{10}, DMU_{12}, and DMU_{13} are in the following ranks, respectively.

5.2. Comparing GFML-NDEA-CI Model with alternative Models

Efficiency measurement is a key business area surrounded by epistemic uncertainty with respect to the dual formed by the object under study and the chosen method to measure performance. Epistemic uncertainty is lack of knowledge on underlying fundamentals or total ignorance of, for example, a possible alternative scenario. This epistemic uncertainty is inherent to the delimitation of the object-method pair under study and manifests itself regardless the identified literature gap, the scale used to measure variables, the variable (input/output) selection and the reproducibility conditions that are intrinsic, to some extent, to the economic sector, industry or DMU set of the application. While the proper identification of a literature gap is relevant for advancing the body of knowledge, especially in DEA-based models, where a plethora of methods are designed to treat specific aspects of a productive network, or in what scales their inputs/outputs are measured, research gaps and measurement scales do not themselves mitigate epistemic uncertainty, only assuring the aspects of internal validity - in light of the current body of knowledge - and scale validity - the proper analytical models were developed to adequately handle the specific nature of what is being measured. Similarly, the choice of economic sector and the selection of key variables are relevant issues for assuring research reproducibility as respect to similar contexts, where generalization of results can be developed with more certainty.

Hence, as regards this paper, epistemic uncertainty can be conceptualized as the scientific uncertainty in the process of modelling. It is due to limited data and knowledge, and possible tools to mitigate epistemic uncertainty in the field of soft computing are the execution of sensitivity analysis, not only by running alternative models, but also using different parameters into the proposed model; and the apprehension of information entropy principles, for improving decision-making as regards whether or not a given model in contributing to reduce epistemic uncertainty. Information entropy is the cornerstone of information theory, providing a constructive criterion for setting up probability distributions of computed scores on the basis of partial knowledge, while enabling a type of statistical inference based on the heterogeneity or dispersion of the scores where no extra biases or uncalled assumptions can enter into the analysis.

Given that in this paper there are at least three distinctive features of our proposed model - serial structure, constant input (composite indicator), and hierarchical structure (multiple-layer) - it would be desirable to find previous papers that, as using composite index measurements, also had a single-layer series network structure. However, to the best of our knowledge, no article with these characteristics has been published so far. Hence, the basic two-stage DEA model of Wanke and Barros [24], after becoming a composite index structure in both deterministic and fuzzy situations were defined as alternative models which used for the sake of comparison using the same dataset presented here. Tables 4 report on the respective scores and entropies for the alternative models considered in the epistemic uncertainty analysis. Note that the Fuzzy CI counterpart of two-stage DEA model of Wanke and Barros [24], i.e. fuzzy NDEA-CI model, is developed by removing the generalized Multiple-Layer property incorporated into the GFML-NDEA-CI model and deterministic two-stage DEA-based CI model, i.e. deterministic NDEA-based CI model, is developed by incorporating *pl*=1 into Fuzzy NDEA-CI model.

Considering Definition 5, the aggregated fuzzy CI score of PBB maturity for three $pl \in \{0, 0.5, 1\}$ for Fuzzy NDEA-CI and GFML-NDEA-CI model are presented in Table 4. In addition, last column of Table 4 indicated the score of deterministic NDEA-based CI model. The numbers in parentheses given in Table 4 are the ranks of the 14 governmental agencies according to their overall efficiency scores for three possibility levels *pl* for the fuzzy NDEA-CI model, multi-layer GFML-NDEA-CI model and deterministic NDEA-based CI model.

Table 4. Comparing the overall efficiency scores of 14 governmental agencies based on the fuzzy one-layer NDEA-CI model and multi-layer GFML-NDEA-CI model.

DMUs	Sector	pl=0		pl=0.5		pl=1		Deterministic CI counterpart of two-stage DEA ([24])
		Fuzzy CI counterpart of two-stage DEA [24]	Proposed model	Fuzzy CI counterpart of two-stage DEA [24]	Proposed model	Fuzzy CI counterpart of two-stage DEA [24]	Proposed model	
DMU_{01}	Banking	0.826 (6)	0.625 (9)	0.902 (6)	0.666 (9)	1 (1)	0.764 (10)	1 (1)
DMU_{02}	Economic planning	0.826 (6)	0.636 (8)	0.901 (7)	0.68 (8)	1 (1)	0.780 (7)	1 (1)
DMU_{03}	Social security	0.768 (10)	0.594 (11)	0.858 (10)	0.656 (11)	0.971 (10)	0.734 (12)	1 (1)
DMU_{04}	Electric energy	0.599 (14)	0.44 (14)	0.681 (14)	0.486 (14)	0.790 (14)	0.590 (14)	0.796 (14)
DMU_{05}	Economic planning	0.832 (1)	0.825 (1)	0.908 (1)	0.903 (1)	1 (1)	1 (1)	1 (1)
DMU_{06}	Economic planning	0.829 (4)	0.642 (7)	0.907 (3)	0.689 (7)	1 (1)	0.787 (5)	1 (1)
DMU_{07}	Banking	0.831 (2)	0.647 (6)	0.908 (1)	0.696 (6)	1 (1)	0.785 (6)	1 (1)
DMU_{08}	Tourism	0.805 (9)	0.672 (5)	0.879 (9)	0.725 (4)	0.979 (9)	0.766 (8)	1 (1)
DMU_{09}	Tourism	0.698 (12)	0.54 (12)	0.795 (12)	0.631 (12)	0.926 (11)	0.765 (9)	0.942 (11)
DMU_{10}	Banking	0.83 (3)	0.82 (2)	0.907 (3)	0.898 (2)	1 (1)	0.988 (2)	1 (1)
DMU_{11}	Social security	0.721 (11)	0.607 (10)	0.798 (11)	0.661 (10)	0.896 (13)	0.753 (11)	0.909 (13)
DMU_{12}	Social security	0.828 (5)	0.713 (3)	0.906 (5)	0.788 (3)	1 (1)	0.882 (3)	1 (1)
DMU_{13}	Banking	0.821 (8)	0.677 (4)	0.90 (8)	0.724 (5)	1 (1)	0.796 (4)	1 (1)
DMU_{14}	Electric energy	0.654 (13)	0.476 (13)	0.752 (13)	0.521 (13)	0.924 (12)	0.626 (13)	0.933 (12)
NED[a]		0	0	0	0	8	1	10
SDS[b]		0.0745	0.105	0.0699	0.1123	0.0589	0.1086	0.0568
DS[c]		0.232	0.385	0.228	0.417	0.21	0.41	0.2039
Max		0.832	0.825	0.908	0.903	1	1	1
Min		0.599	0.440	0.681	0.486	0.79	0.59	0.7961
Entropy index		0.9982	0.9948	0.9987	0.9950	0.9993	0.9964	0.9993
Mean		0.7763	0.6367	0.8573	0.6946	0.9633	0.7869	0.97

[a] NED: The number of efficient DMUs; [b] SDS: Standard deviation of scores; [c] DS: Domain of scores

Table 4 indicates that the discriminative power of the GFML-NDEA-CI model is stronger than that of its one-layer counterpart. The number of efficient DMUs (NED), the standard deviation of scores (SDS), domain of scores (DS) of DMUs, and information entropy are four main measures presented to prove the above claim.

The first measure is NED. The results at $pl = 1$ indicate that NED in the fuzzy two-stage DEA and deterministic two-stage DEA model are 8 and 10, respectively. This means that due to the large number of sub-indicators, the one-layer models was not able to discriminate among DMUs, while in the multi-layer model, only one DMU is efficient and the rest are inefficient, which is due to the hierarchical structure of the indicators, as well as imposing upper and lower bounds on the weights of each sub-indicator.

The SDS among DMUs is the second measure to evaluate the discriminative power of the model such that when the SDS is higher, it means that the model has been able to make a greater discrimination among DMUs. As shown in Table 4, the SDS of the multi-layer model at all levels of pl is higher than that of the one-layer model. For instance, for pl =1, the SDS in the multi-layer model is 10.86 while it is 5.89 in the one-layer model.

The third measure is the DS, which is the difference between the maximum and minimum scores. Accordingly, the DS of the two models indicates that the discriminative power of the multi-layer model is more than that of the one-layer model. For example, for $pl = 1$, the DS of the one-layer model is 21, while it is 41 in the multi-layer model. This point is indicated as a graphical comparison. As shown in Figure 5, the score dispersion of the multi-layer model is higher than that of the one-layer, which indicates the higher resolution of the multilayer model as the highest score of both models (for *pl=1*) are equal to 100% and the lowest score for the multi-layer and one-layer are 59% and 79%, respectively. The following figure also indicates the density of DMU scores in the one-layer model are around 1, while for the multiple layer they are between [0.7, 0.8].

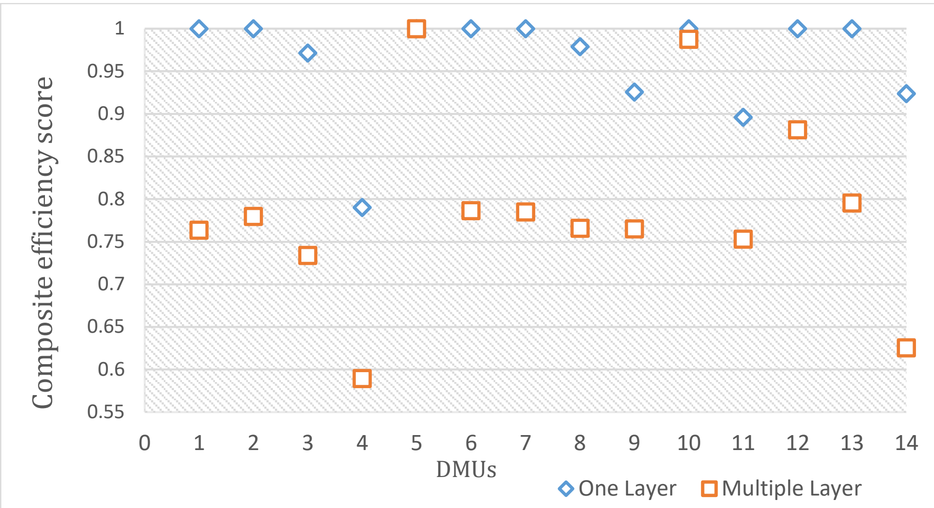


Figure 5. The spread of composite efficiency scores among DMUs for $pl = 1$ in one-layer vs. multi-layer models

The fourth measure is information entropy, which is an important metric of information theory discipline and captures uncertainty in terms of the distributional shape rather than in terms of the probability mass concentration around the mean, which are captured by statistical moments. Lower information entropy are preferable [36]. Depending on the entropy characteristics, the randomness and dispersion of a criterion/construct can be determined by calculating the information entropy. In this article, information entropy is used to analyze the distribution of the CI score (efficiency score). As shown in Table 4, the efficiency scores obtained using the developed multi-layer models for all three

pl are distinct from their respective deterministic two-stage DEA and fuzzy two-stage NDEA models in terms of information entropy.

Comparing the rankings obtained from the GFML-NDEA-CI model with those obtained from the conventional fuzzy NDEA-CI model, there are a couple of points that are worth discussing briefly.

Spearman's correlation coefficient is a measure of the linear association between two variables that are available in the ordinal scale. That is, this coefficient measures the strength of the association between two ranked variables [37,38]. Based on the Spearman rank correlation coefficient test with the critical statistic value of 0.645 for a significance level α=0.01, the rankings of the 14 agencies obtained from these two alternative models are not significantly different for all three pessimistic, indifferent, and optimistic situations with different possibility levels, pl. Applying the Spearman rank correlation test, the test statistics r_s (and the corresponding p-values) for the GFML-NDEA-CI model and conventional fuzzy NDEA-CI model are 0.840 ($p < 0.001$), 0.813 ($p < 0.001$), and 0.825 ($p < 0.001$) for possibility levels $pl = 0$, $pl = 0.5$, and $pl = 1$, respectively.

By examining the aggregated fuzzy efficiency scores obtained from both models, it is recognized that the discrimination power of GFML-NDEA-CI model is much higher than that of the conventional fuzzy NDEA-CI model. This realization indicates a merit of the GFML-NDEA-CI model that it reports more informative results and more accurate rankings.

6. Managerial Implications

The managerial implications of this research are threefold. First of all, a comprehensive multiple-layer NDEA model is provided to decision-makers for structuring different indicators in a flexible hierarchic framework. While these indicators can be organized in series or in parallel, score discrimination remains high in comparison to single-layer models as long as the multiple hierarchic layers help in mitigating the curse of dimensionality imposed by small number of DMUs in comparison to a large number of indicators. Secondly, as long as indicator weights are determined endogenously by the MLNDEA model in a fashion that the overall performance of each DMU is attained, possible improvement paths in the budgetary process can be derived specifically for each unit, drawing the decision-maker's attention to the capability weakness that should be improved and result strengths that should be replicated in different units. Third, the use of fuzzy numbers as a means for handling the surrounding vagueness with respect to indicator measurement and overall knowledge of decision-makers with respect to environment allows drawing different performance scenarios that could be explored by decision-makers in terms of minimal performance requirements (pessimistic) or superior performance achievement (optimistic).

Figure 7 presents the average aggregated efficiency scores of 14 agencies, which are grouped into five sectors including tourism, electric energy, banking, social security, and economic planning.

Results obtained from a small number of agencies working towards establishing and implementing a PBB system revealed that relevant maturity is the highest in agencies involved in the economic planning (CI score of 0.757) and banking (CI score of 0.746) sectors, respectively. The CI scores for agencies involved in social security and tourism are 0.702 and 0.678, respectively, which are close to the average CI score for all agencies, i.e., 0.690 (presented by the dotted line). Agencies involved in the electric energy sector have a CI score of 0.504, which is much lower than the average CI score.

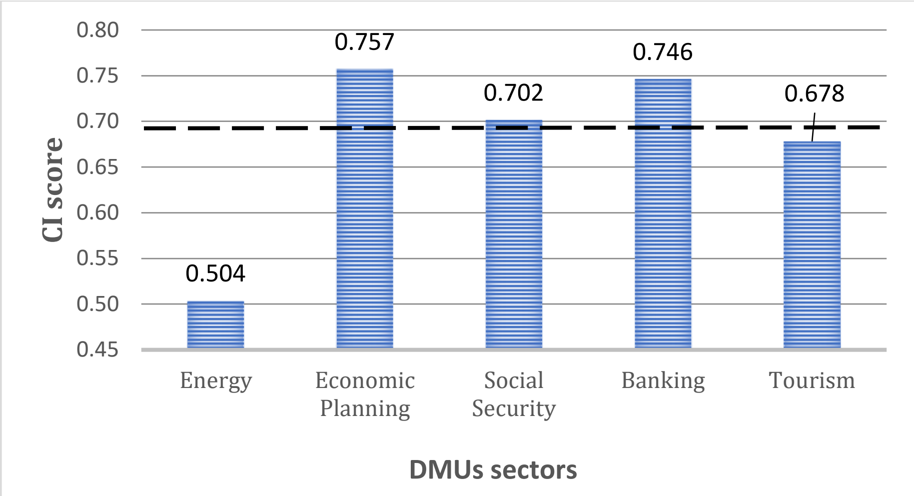


Figure 6. Average aggregated efficiency scores for five different sectors

As noted by [12], the weights to each indicator can be interpreted as the importance of that indicator for the corresponding DMU score. Looking at the assigned weights (shares) provided in the Supplementary Material, it is possible to note that all 14 agencies are in a poor situation in three indicators including "information management", "performance management", and "control & monitoring" due to receiving the lowest weights, and also all 14 agencies are in a strong position in "planning" due to receiving the highest weights. The weights assigned to the other capability indicators including "costing system" and "process management & documentation" vary among DMUs. For instance, in the banking agencies ($\mathrm{DMU}_{01}, \mathrm{DMU}_{07}$, and DMU_{13}), after "planning", the "costing system" has been considered more important than other indicators. However, in economic planning agencies (DMU_{02}, DMU_{05}, and DMU_{06}) and social security agencies (DMU_{03}, DMU_{11} and DMU_{12}), after “planning”, their performance in the "process management & documentation" were better ranked than the other indicators.

These results appear to be consistent with empirical evidence. Being ahead in planning could be interpreted that even though many years have passed since the PBO of Iran implemented the PBB system in 2013, only very recently has this planning activity received serious attention. Such recent attention emanates from the fact that in order to renew a budget agreement, the PBO requires all agencies to redefine the planning subsystem in accordance with the PBB system's executive guidelines, otherwise the PBO will not allocate funds to them. Also, according to budget experts in Iran, one of the most important challenges faced in establishing the PBB system in Iran is the lack of comprehensive information systems in many agencies together with the existence of dispersed financial and non-financial indicators.

Similar to information management, "performance management" and "control & monitoring" have also received less attention. Apart from organizational ability, one of the reasons for paying less attention to these two capability indicators can be traced back to the assessment focus adopted by the SAC of Iran. In recent years, the SAC has focused more on financial audit and less on performance audit, therefore different organizations were not required to pay attention to the performance dimension of the budgeting system. However, since 2018 the SAC's approach has been revised from financial auditing to both financial and performance auditing. This has driven organizations to pay more attention to the performance aspect of their budget.

With the same approach, the weights assigned to "result" indicators show that most agencies are in a strong position with respect to "budget discipline" and most agencies are in a poor situation in "applying performance information". The weights assigned to the two other result indicators including "transparency & accountability" and "applying costing information" vary among DMUs. Being ahead in “budget discipline”, in addition to its many organizational benefits, can be attributed to the regulatory role of organizations such as the SAC. As noted earlier, as the SAC focuses on compliance with laws and regulations (i.e., financial audit) and less on performance audit, the "applying performance information" indicator is less favored by organizations.

7. Conclusion

To analyze a large amount of information, it is easier to look at the different indicators as an integrated construction than to find a specific pattern to analyze the indicators separately. Composite indicators are increasingly considered as a valuable tool for policy analysis and public communication. Therefore, in this study we investigated applying the two-stage data envelopment analysis for building composite indicators.

A number of practical implications can be derived. First, when the modeler is faced with a large number of indicators in different categories, instead of aggregating or ignoring some indicators, applying a multi-layered approach can achieve more reliable results (i.e. lower information entropy). Note that the high similarity of the ranking result based on these two models verified its robustness. Secondly, non-parametric optimization was implemented upon a comprehensive Multiple-Layer hierarchic structure, thus providing an alternative model to pairwise comparisons in some other comparison methods (e.g. Analytic Network Process[1] or even to the statistical modeling within the ambit of Structural Equation Modeling[2]). Thirdly, applying a multilayer DEA approach through intra-category weighting represents important content about strong and poor status of indicators in each DMU, which provides insights into ways to improve the system's efficiency. Fourthly, applying fuzzy logic to deal with uncertain data on some indicators and sub-indicators allows the desired flexibility to handle indicators of different scales such as metric, ordinal, or even nominal, as well as when decisions are made under uncertainty/vagueness, such as in the case of PBB in Iran.

Further research venues could consider not only different approaches for modeling data uncertainty such as Z-numbers or interval data, but also data randomness by incorporating stochastic performance elements over the course of time by adding a temporal dimension in the multiple-layer network DEA model via links and carry-overs. Two-Dimensional Fuzzy-Monte Carlo approaches could be employed to identify which performance components are most subject to vagueness to the detriment of randomness, and vice-versa, as capabilities and results may be influenced differently from these different uncertainty sources. Improvement programs could be designed in terms of quality control tools for those indicators subjected the most to randomness, while online or real-time database capture and measurement could be implemented for those indicators subject to vagueness, for instance. In addition, incorporating the multiple-layer concept into other Network DEA models is worthwhile.

References

[1] M.J. Farrell, The measurement of productive efficiency, J. R. Stat. Soc. Ser. A. 120 (1957) 253–281.

[2] C. Kao, Network data envelopment analysis: A review, Eur. J. Oper. Res. 239 (2014) 1–16.

[3] C. Kao, Network data envelopment analysis, Int. Ser. Oper. Res. Manag. Sci. 10 (2017) 973–978.

[4] H. Fukuyama, R. Matousek, Modelling bank performance: A network DEA approach, Eur. J. Oper. Res. 259 (2017) 721–732.

[5] P. Wanke, M.A.K. Azad, A. Emrouznejad, J. Antunes, A dynamic network DEA model for accounting and financial indicators: A case of efficiency in MENA banking, Int. Rev. Econ. Financ. 61 (2019) 52–68.

[6] K. Wang, W. Huang, J. Wu, Y.N. Liu, Efficiency measures of the Chinese commercial banking system using an additive two-stage DEA, Omega. 44 (2014) 5–20.

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

[8] A. Hatami-Marbini, S. Saati, Efficiency evaluation in two-stage data envelopment analysis under a fuzzy environment: A common-weights approach, Appl. Soft Comput. 72 (2018) 156–165.

[9] M. Duygun, D. Prior, M. Shaban, E. Tortosa-Ausina, Disentangling the European airlines efficiency puzzle: A network data envelopment analysis approach, Omega (United Kingdom).

[1] - Analytic Network Process (ANP) provides an approach for feeding judgments and measurements to properly compute (ratios scale) priorities to distribute their impact or influence over the factors and groups of factors in the decision. For more details see [39].

[2] - Structural Equation Modeling (SEM) is a powerful technic that can combine complex path models with latent variable (factors). Using SEM, researchers can specify confirmatory factor analysis model and complex path models. For more details see [40].

60 (2016) 2–14.
[10] E. Soltanzadeh, H. Omrani, Dynamic network data envelopment analysis model with fuzzy inputs and outputs: An application for Iranian Airlines, Appl. Soft Comput. 63 (2018) 268–288.
[11] H. Omrani, K. Shafaat, A. Emrouznejad, An integrated fuzzy clustering cooperative game data envelopment analysis model with application in hospital efficiency, Expert Syst. Appl. 114 (2018) 615–628. http://www.sciencedirect.com/science/article/pii/S0957417418304986.
[12] Y. Shen, E. Hermans, D. Ruan, G. Wets, T. Brijs, K. Vanhoof, A generalized multiple layer data envelopment analysis model for hierarchical structure assessment: A case study in road safety performance evaluation, Expert Syst. Appl. 38 (2011) 15262–15272.
[13] S.M. Mirhedayatian, M. Azadi, R. Farzipoor Saen, A novel network data envelopment analysis model for evaluating green supply chain management, Int. J. Prod. Econ. 147 (2014) 544–554.
[14] T. Badiezadeh, R.F. Saen, T. Samavati, Assessing sustainability of supply chains by double frontier network DEA: A big data approach, Comput. Oper. Res. 98 (2018) 284–290.
[15] X. Zhou, W. Pedrycz, Y. Kuang, Z. Zhang, Type-2 fuzzy multi-objective DEA model: An application to sustainable supplier evaluation, Appl. Soft Comput. J. 46 (2016) 424–440.
[16] P. Yin, J. Chu, J. Wu, J. Ding, M. Yang, Y. Wang, A DEA-based two-stage network approach for hotel performance analysis: An internal cooperation perspective, Omega. (2019).
[17] M.R. Amini, A. Azar, M.D. Nayeri, K. Bayat, Developing a performance-based budgeting maturity model and constructing a DEA-based composite indicator to measure it ’ s score, Ind. Eng. Manag. Syst. 18 (2019) 144–154.
[18] C. Huang, C. Dai, M. Guo, A hybrid approach using two-level DEA for financial failure prediction and integrated SE-DEA and GCA for indicators selection, Appl. Math. Comput. 251 (2015) 431–441.
[19] W. Meng, D. Zhang, L. Qi, W. Liu, Two-level DEA approaches in research evaluation, Omega. 36 (2008) 950–957.
[20] C. Huang, F.N. Ho, Y. Chiu, Measurement of tourist hotels′ productive efficiency, occupancy, and catering service effectiveness using a modified two-stage DEA model in Taiwan, Omega. 48 (2014) 49–59.
[21] L. Cherchye, W. Moesen, N. Rogge, T. Van Puyenbroeck, An introduction to “benefit of the doubt” composite indicators, Soc. Indic. Res. 82 (2007) 111–145.
[22] Y. Shen, E. Hermans, T. Brijs, G. Wets, Fuzzy data envelopment analysis in composite indicator construction, Stud. Fuzziness Soft Comput. (2014) 89–100.
[23] M.Z. Mahmoudabadi, A. Azar, A. Emrouznejad, A novel multilevel network slacks-based measure with an application in electric utility companies, Energy. 158 (2018) 1120–1129. http://www.sciencedirect.com/science/article/pii/S0360544218310053.
[24] P. Wanke, C. Barros, Two-stage DEA: An application to major Brazilian banks, Expert Syst. Appl. 41 (2014) 2337–2344. https://www.sciencedirect.com/science/article/pii/S0957417413007653 (accessed April 29, 2019).
[25] C. Kao, A linear formulation of the two-level DEA model, Omega. 36 (2008) 958–962.
[26] R. Ezzati, E. Khorram, R. Enayati, A new algorithm to solve fully fuzzy linear programming problems using the MOLP problem, Appl. Math. Model. 39 (2015) 3183–3193. http://www.sciencedirect.com/science/article/pii/S0307904X13001698.
[27] J. Pourmahmoud, N. Bafekr Sharak, Measuring Cost Efficiency with New Fuzzy DEA Models, Int. J. Fuzzy Syst. 20 (2018) 155–162.
[28] Y. Shen, D. Ruan, E. Hermans, T. Brijs, G. Wets, K. Vanhoof, Modeling qualitative data in data envelopment analysis for composite indicators, Int. J. Syst. Assur. Eng. Manag. 2 (2011) 21–30.
[29] P. Guo, H. Tanaka, Fuzzy DEA: A perceptual evaluation method, Fuzzy Sets Syst. 119 (2001) 149–160.
[30] P. Guo, Fuzzy data envelopment analysis and its application to location problems, Inf. Sci. (Ny). 179 (2009) 820–829.
[31] R.G. Dyson, E. Thanassoulis, Reducing weight flexibility in data envelopment analysis, J. Oper. Res. Soc. 39 (1988) 563–576.

[32] J.B. Gilmour, D.E. Lewis, Does performance budgeting work? An examination of the office of management and budget's PART scores, Public Adm. Rev. 66 (2006) 742–752.
[33] S.G. Mauro, Public sector performance-based budgeting in Italy, in: Performance-Based Budg. Public Sect., Palgrave Macmillan, Cham, 2019: pp. 125–140.
[34] M. Freudenberger, Composite indicators of country performance: a critical assessment, OECD Sci. Technol. Ind. Work. Pap. (2003) 1–30.
[35] OECD, Handbook on constructing composite indicators, OECD Publishing, 2008.
[36] J. Gerami, M.R. Mozaffari, P.F. Wanke, A Multi-Criteria Ratio-Based Approach for Two-Stage Data Envelopment Analysis, Expert Syst. Appl. (2020) 113508.
[37] C. Kao, S.T. Liu, Cross efficiency measurement and decomposition in two basic network systems, Omega (United Kingdom). 83 (2019) 70–79.
[38] D.U.A. Galagedera, I. Roshdi, H. Fukuyama, J. Zhu, A new network DEA model for mutual fund performance appraisal: An application to U.S. equity mutual funds, Omega (United Kingdom). 77 (2018) 168–179.
[39] T.L. Saaty, Decision making with dependence and feedback: The analytic network process, RWS Publ., 1996.
[40] J.J. Hox, T.M. Bechger, An introduction to structural equation modeling, Fam. Sci. Rev. 11 (1998) 354–373.