Expert Systems With Applications 195 (2022) 116571

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Analysis of the impact of DMUs on the overall efficiency in the event of a merger

A. Saavedra-Nieves [a,*], M.G. Fiestras-Janeiro [b]

[a] Departamento de Estatística, Análise Matemática e Optimización, Universidade de Santiago de Compostela, Santiago de Compostela 15705, Spain
[b] Departamento de Estatística e Investigación Operativa, Universidade de Vigo, Vigo 36271, Spain

ARTICLE INFO

Keywords:
DEA model
Multiagent system
Partition function form game
Ranking method
Sampling

ABSTRACT

This paper addresses several mechanisms for overall ranking Decision Making Units (DMUs) according to the contribution of DMUs to the relative efficiency score of a merger considering aggregate units The possible organization of agents outside each possible merger naturally influences the relative efficiency score, which motivates the use of games in partition function form and specific ranking indices for DMUs based on the Shapley value. Several computational problems arise in their exact computation when the number of DMUs increases. We describe two sampling alternatives to reduce these drawbacks. Finally, we apply these methods to analyse the efficiency of the hotel industry in Spain.

1. Introduction

The analysis of expert or intelligent multiagent situations cannot be overlooked because the global interconnections among their components is a key factor of society. Such relations ensure that information about the decisions in a group of agents is influenced by the outcomes of other groups of agents. In a collaborative scheme, considering the coalitional externalities of agents in the game-theoretical framework used for modelling such cooperation can derive a much more realistic fitting of the used transferable utility (TU) game approach. For instance, resource allocation problems, which may be influenced by factors such as environmental or climate change influence, or Data Envelopment Analysis are clear examples of the above.

Data Envelopment Analysis (DEA) is a nonparametric methodology in operations research and economics, introduced by Charnes et al. (1978), and it is used to assess the relative efficiency score of a set of homogeneous Decision Making Units (DMUs). These DMUs are characterized by the use of certain inputs to produce certain outputs. The idea behind these techniques is the estimation of efficient frontiers of the Production Possibility Set (PPS), on which each DMU is projected. Formally, the efficient frontier is given by those nondominated operation points. Thus, a DMU is said to be inefficient if an operation point exists that produces the same outputs using less inputs, or produces more outputs with the same inputs. Immediately, DMUs can be rated as efficient or as inefficient. Although these techniques were proposed for evaluating activities of nonprofit entities, multiple references in the literature have applied them to a wide variety of real problems. We mention (Hua et al., 2007), that apply DEA for ecological efficiency evaluations of factories, and Cooper et al. (2009), that evaluate how effective basketball players are by using DEA. In addition, Yin et al. (2020) evaluate the hotel performance in relation to the impact of operations and marketing capabilities. and Yazdani et al. (2020) use DEA to select the location of logistics centres in Spanish autonomous regions.

The natural collaboration of DMUs within a DEA framework favours the use of cooperative game theory for its analysis in recent years. For instance, Cook and Kress (1999) first use DEA for fixed cost allocation. Lozano (2012) analyses the gains achieved by sharing information on the consumption and the production of DMUs. Yang and Zhang (2015) allocate resources by using a modified Shapley value. An et al. (2019) explore the payoff allocation problem in a three-stage system and Chu et al. (2020) allocate a fixed cost among DMUs with two-stage structures in DEA. As an immediate application, rankings in DEA settings are naturally justified. Li and Liang (2010) use the Shapley value to rank variables according to the efficiencies of DMUs. Ranking DMUs using cooperative game theory has received considerable attention in the literature, since that this process substantially increases the discrimination power of the DEA method. We refer to Lozano et al. (2016) as a comprehensive survey on this topic, although there are other references in the literature focusing on this topic. Moreover, we refer to Li et al. (2016), who uses cooperative game theory to rank efficient DMUs from their efficiencies under DEA. Hinojosa et al. (2017) consider the relative efficiency score of inefficient DMUs when a group of efficient units is dropped from the sample. Omrani et al. (2018)

* Corresponding author.
E-mail addresses: alejandro.saavedra.nieves@usc.es (A. Saavedra-Nieves), fiestras@uvigo.es (M.G. Fiestras-Janeiro).

https://doi.org/10.1016/j.eswa.2022.116571
Received 19 February 2021; Received in revised form 27 September 2021; Accepted 17 January 2022
Available online 5 February 2022
0957-4174/

introduce the class of Integrated Fuzzy Clustering cooperative games DEA, and use the core and the Shapley value for ranking efficient DMUs. In this sense, Demetrescu et al. (2019) also use the Shapley value as a ranking mechanism under DEA for the Italian system of evaluating research quality.

In this paper, we analyse the impact of DMUs on the overall relative efficiency score in the event of a merger from a game theoretical approach. We will consider the DEA formulation in Charnes et al. (1978) due to the vast collection of papers that have proven its reliability in practice and because it is often presented as a basis for many of the formulations and approaches in DEA literature. The first novelty of our proposed approach is that the set of agents considered is now given by the whole set of DMUs, regardless of whether they are individually efficient (or not) under DEA. Moreover, we build new DMUs, called artificial DMUs, which are associated with any group or coalition of our primary DMUs by aggregating the inputs and the outputs of their members. Each artificial DMU can be seen as an entity that collects all inputs and outputs of its members as in Hosseinzadeh Lotfi et al. (2011) or Kritikos (2017), among others. We measure the relative efficiency score of each artificial DMU under any possible configuration of the outsiders. The main innovation lies in the fact that the relative efficiency score of each artificial DMU is obtained under any possible configuration of the remaining DMUs. This perspective allows us to use the model of *partition function form games* (Thrall & Lucas, 1963) to represent these situations. As in the case of cooperative games with transferable utility (TU game), one main goal for partition function form games is to split the global earning in a reasonable way, such as by extending the Shapley value (Shapley, 1953). Papers related to the application of partition function form games have been published. For instance, Pintassilgo and Lindroos (2008) model the management of the North Atlantic bluefin tuna fishery; Liu et al. (2016) model the harvesting costs of the Norwegian springspawning herring; Csercsik and Kóczy (2017) analyse an electrical network to obtain balancing groups of producers and consumers; and Yang et al. (2019) analysed sequencing situations with externalities. Another application is the coalition formation process (see Mahdiraji et al., 2021). In this paper we will use partition form games and the values defined in Albizuri et al. (2005), de Clippel and Serrano (2008), Hu and Yang (2010) and McQuillin (2009) to rank the whole set of DMUs based on the amount allocated to each of them.

Certain computational problems, such as those that motivated Castro et al. (2009) or Maleki (2015), also maintain now. Specifically, for every coalition of DMUs, we need to assign the worth of the cooperation among its members for all possibilities of grouping outside the coalition. Thus, new difficulties arise because when forming a coalition, all possibilities of organization of the remaining agents cannot be evaluated in polynomial time. To solve this new drawback, sampling methodologies (Cochran, 2007) are considered as alternatives.

This paper is organized as follows. Section 2 introduces some basic notions on DEA problems. Section 3 innovatively introduces the class of DEA sum games under the presence of externalities and studies some properties that they satisfy. Section 4 provides several alternatives for ranking DMUs in this new setting based on values for games in partition function form. Section 5 describes two sampling proposals for their estimation in contexts that involve a large number of DMUs. Section 6 illustrates this methodology using data from the hotel industry in Spain. Some concluding remarks are presented in Section 7.

2. On DEA problems

Data environment analysis (DEA) deals with measuring the relative efficiency score of a homogeneous group of *decision-making units* (*DMUs*), which all present a common goal to be achieved. Each DMU uses certain inputs to produce certain outputs. It is noteworthy that these inputs and outputs are usually multiple in character and may also assume a variety of forms, which permits only ordinal measurements. Charnes et al. (1978) initially analyse this problem from a mathematical point of view. This contribution has high relevance in economic environments due to the large number of real applications.

Let $N = \{1, \dots, n\}$ be the system of DMUs under study. Each DMU is characterized in a DEA problem by the elements that are enumerated below.

- There exist m different inputs. The value x_{ki} denotes the amount of input k, with $k = 1, \dots, m$ used by DMU i for every $i \in N$. All this information can be grouped in the subvectors X_i, which contain the input information for every DMU i. If they are concatenated, they directly determine the input matrix $X = (X_1, \dots, X_n)$.
- There exist l different outputs. In this case, y_{ki} denotes the amount of output k, with $k = 1, \dots, l$ produced by DMU i for every $i \in N$. Analogously, we can organize this information in the subvector Y_i that includes the output information for DMU i. In this case, the matrix $Y = (Y_1, \dots, Y_n)$ contains the output data for all the DMUs in N.

In this paper, we consider the output-oriented DEA model with constant returns to scale that was formally described in Charnes et al. (1978). However, the presented procedures can be applied for any DEA model. Adler et al. (2002) note that the output-oriented and the input-oriented DEA models of Charnes et al. (1978) lead to the same results. However, this consistency is not observed for the variable returns to scale model considered in Banker et al. (1984). Even so, Problem FP can still be used to determine returns to scale by using the model of Charnes et al. (1978) in envelopment form as in Banker et al. (2011).

The performance of a DEA problem is based on the following premises. Formally, a DEA problem enables a global evaluation of the individual efficiency score of a DMU in a certain system relative to all other DMUs. This value, which is usually known in the literature as the *relative efficiency score* of a DMU, can be defined in terms of the profitable character of its operation. More specifically, this profit is defined in terms of the maximization of the ratio of its weighted outputs with respect to its weighted inputs, which is subject to the condition that similar ratios for every DMU are less than or equal to unity. Taking into account these assumptions, this analysis is formulated through a mathematical programming problem that models this class of situations. This problem corresponds to a *multiagent DEA problem* that will be denoted by $(N; X; Y)$.

Let i_0 be a certain DMU. The relative efficiency score of DMU i_0, according to the output-oriented DEA model with constant returns to scale, is obtained by maximizing the objective function in the optimization problem formulated below:

FP

$$\begin{aligned} \max \quad & \theta_{i_0} = \frac{\sum_{r=1}^{l} u_r y_{r0}}{\sum_{j=1}^{m} v_j x_{j0}} \\ \text{subject to} \quad & \\ & \frac{\sum_{r=1}^{l} u_r y_{ri}}{\sum_{j=1}^{m} v_j x_{ji}} \le 1, \forall i = 1, \dots, n. \\ & u_r, v_j \ge 0; \; r = 1, \dots, l; \; j = 1, \dots, m. \end{aligned} \tag{1}$$

More specifically, the relative efficiency score of DMU i_0 is given by the optimal value of the quotient θ_{i_0}, $\theta^*_{i_0} = \frac{\sum_{r=1}^{l} u^*_r y_{r0}}{\sum_{j=1}^{m} v^*_j x_{j0}}$, with $v^* = (v^*_1, v^*_2, \dots, v^*_m)$ and $u^* = (u^*_1, u^*_2, \dots, u^*_l)$ being the optimal solution of the problem. By satisfying also feasibility, it always holds that $\theta^*_{i_0} \le 1$.

Since the problem is a fractional programming problem, it is necessary to apply some transformations that reduce the difficulty of reaching a solution. From Charnes et al. (1978), the following formulation is introduced in terms of a linear programming problem, which is an alternative to Problem FP, for measuring the relative efficiency score of DMU i_0.

MP

$$\begin{array}{ll} \max & \eta_{i_0} \\ \text{subject to} & \\ & -\sum_{i\in N} y_{ri}\lambda_i + y_{ri_0}\eta_{i_0} \le 0,\ r=1,\dots,l \\ & \sum_{i\in N} x_{ji}\lambda_i \le x_{ji_0},\ j=1,\dots,m \\ & \lambda_i \ge 0,\ \forall i \in \{1,\dots,n\} \\ & \eta_{i_0} \in \mathbb{R}. \end{array} \tag{2}$$

In this case, the optimal solution is given by $\lambda^* = (\lambda_1^*, \lambda_2^*, \dots, \lambda_n^*)$ and $\eta_{i_0}^*$. We can also consider the dual problem of (2), whose formulation is given below:

DP

$$\begin{array}{ll} \min & \rho_{i_0} = \sum_{j=1}^{m} \omega_j x_{ji_0} \\ \text{subject to} & \\ & -\sum_{r=1}^{l} \mu_r y_{ri} + \sum_{j=1}^{m} \omega_j x_{ji} \ge 0, \forall i \in \{1,\dots,n\}. \\ & \sum_{r=1}^{l} \mu_r y_{ri_0} = 1 \\ & \mu_r, \omega_j \ge 0;\ r=1,\dots,l;\ j=1,\dots,m. \end{array} \tag{3}$$

The optimal value of the problems described above, which is given by $\theta_{i_0}^*$, $\eta_{i_0}^*$ and $\rho_{i_0}^*$, satisfies that

$$\eta_{i_0}^* = \rho_{i_0}^* \text{ and } \theta_{i_0}^* = \frac{1}{\eta_{i_0}^*}.$$

Taking into account the previous relations on the relative efficiency scores, if DMU i_0 is efficient, then Problem FP ensures $\theta_{i_0}^* = 1$; thus, $\rho_{i_0}^* = 1$. However, $\theta_{i_0}^* = 1$ does not ensure the efficiency of DMU i_0. Moreover, using less inputs or producing more outputs (with the amounts of outputs or inputs remaining constant, respectively) does not affect the worth of $\theta_{i_0}^*$. The DMU resulting from such inputs and outputs would be qualitatively more efficient than the initial one. For additional details, see Charnes et al. (1978). We only assess the relative efficiency score of a DMU through the objective function of Problem MP, for example, regardless of whether the DMU is efficient (or not).

3. On the cooperation of DMUs under externalities

In this section, we analyse the problem of cooperation in DEA using a different approach from the usual collaboration concepts such as those in Hinojosa et al. (2017), Li and Liang (2010) and Lozano et al. (2016). The overall relative efficiency score of a merger of a group of DMUs is influenced by the organization of the remaining DMUs and their possible mergers. Thrall and Lucas (1963) introduced the model of *games in partition function form* to describe situations such as these, in which the worth of a coalition substantially depends on how the remaining agents are organized. In this framework the basic organization of agents is called an *embedded coalition*, which is a pair whose first component is an element of a partition and whose second component contains the remaining elements of the partition (sometimes, the whole partition).

Let N be a set of agents and $\Pi(N)$ be the set of partitions of N. Formally, an *embedded coalition* is given by a pair $(S;P)$ with $S \subseteq N$ and $P \in \Pi(N \setminus S)$.[1] The empty set always belongs to any partition $P \in \Pi(N)$. Nevertheless, we consider embedded coalitions of type $(\emptyset;P)$, for every $P \in \Pi(N)$. The family of those nonempty embedded coalitions is denoted by

$$EC^N = \{(S;P) :\ P \in \Pi(N \setminus S) \text{ and } S \subseteq N\}. \tag{4}$$

[1] We use this notation for simplicity. If $S = N$, then we use $(N;\emptyset)$.

A *partition function form game* or a *game with externalities* is formally defined by a function $e : EC^N \longrightarrow \mathbb{R}$ such that $e(\emptyset;P) = 0$. $PG(N)$ denotes the family of all games with externalities with player set N.

Consider the multiagent DEA problem $(N;X;Y)$. In this context, measuring the relative efficiency score of DMUs under cooperation varies when embedded coalitions are considered. If $S \subseteq N$ is formed, then an artificial DMU $[i_S]$ is defined. This artificial DMU makes use of a certain amount of inputs to produce another amount of outputs, with both quantities obtained from the original values. Different proposals associated with this aim has been reported. For simplicity, we consider the natural case of the direct aggregation, such as in Hosseinzadeh Lotfi et al. (2011) and Kritikos (2017). Formally, if DMUs in $S \subseteq N$ decide to cooperate, then the *sum-sum artificial DMU* $[i_S]$ is defined by an input vector $x_{[i_S]} = (x_{1[i_S]}, \dots, x_{m[i_S]})$ and an output vector $y_{[i_S]} = (y_{1[i_S]}, \dots, y_{l[i_S]})$ given by

$$\begin{aligned} x_{k[i_S]} &= \sum_{j\in S} x_{kj}, \text{ with } k = 1,\dots,m, \\ y_{k[i_S]} &= \sum_{j\in S} y_{kj}, \text{ with } k = 1,\dots,l. \end{aligned} \tag{5}$$

As mentioned, the organization of agents in $N \setminus S$ according to a coalition structure that describes their cooperation affinities also affects the relative efficiency score of $[i_S]$. Below, we formally describe its performance. Let $P \in \Pi(N)$ be a coalition structure for the agents in N. We define the artificial DMU $[i_S]$ for every $S \in P$, in such way the new set of DMUs is given by $N^P = \{[i_S] :\ S \in P\}$, with $x_{[i_S]}$ and $y_{[i_S]}$ representing the corresponding input and output vectors obtained from (5) for each $S \in P$, respectively. Thus, the resulting *DEA problem* will be denoted by $(N^P; X^P; Y^P)$.

It is possible to define a game with externalities associated with any multiagent DEA problem $(N;X;Y)$.

Definition 3.1. Let $(N;X;Y)$ be a multiagent DEA problem. We can define an associated *partition function form game* $(N;X;Y;\mathbf{e})$ as follows:

$$\mathbf{e}(S;P) = \begin{cases} \dfrac{1}{\eta^*_{[i_S],P}}, & \text{if } \emptyset \ne S \subseteq N,\ P \in \Pi(N \setminus S), \\ 0, & \text{otherwise.} \end{cases}$$

$\eta^*_{[i_S],P}$ is the optimal value of Problem (2) for the DEA problem $(N^{P\cup\lceil S\rceil}; X^{P\cup\lceil S\rceil}; Y^{P\cup\lceil S\rceil})$, with $P \cup \lceil S\rceil \in \Pi(N)$ having S as a block,[2] for DMU $i_0 = [i_S]$. Thus, $(N;X;Y;\mathbf{e})$ is called a *DEA sum game*, and for notational convenience, it will be denoted by $\mathbf{e}$.

Specifically, by considering the output-oriented DEA model with constant returns to scale, the optimal value of the DMU i in Problem (2) determines the value of $\mathbf{e}(\{i\};\{\{j\} :\ j \in N\setminus\{i\}\})$. In the remainder, $\lfloor S\rfloor$ denotes $\{\{j\} :\ j \in S\}$ for each $S \subseteq N$. Under externalities, given a nonempty coalition S of N and $P \in \Pi(N\setminus S)$, Problem (2) is rewritten for any $(N^{P\cup\lceil S\rceil}; X^{P\cup\lceil S\rceil}; Y^{P\cup\lceil S\rceil})$ and any $i_0 = [i_S]$ as follows:

$\mathbf{MP(S;P)}$

$$\begin{array}{ll} \max & \eta_{[i_S],P} \\ \text{subject to} & \\ & -\sum_{T\in P\cup\lceil S\rceil} y_{r[i_T]}\lambda_{[i_T]} + y_{ri_S}\eta_{[i_S],P} \le 0, \\ & r = 1,\dots,l, \\ & \sum_{T\in P\cup\lceil S\rceil} x_{j[i_T]}\lambda_{[i_T]} \le x_{j[i_S]},\ j=1,\dots,m \\ & \lambda_{[i_T]} \ge 0,\ T \in P \cup \lceil S\rceil \\ & \eta_{[i_S],P} \in \mathbb{R}. \end{array} \tag{6}$$

[2] $\lceil S\rceil$ means that S is an element of the partition.

Equivalently, the dual problem of MP(S;P) is given by

DP(S; P)

$$\begin{aligned} \min \quad & \rho_{[i_S],P} = \sum_{j=1}^{m} \omega_j x_{j[i_S]} \\ \text{subject to} \quad & -\sum_{r=1}^{l} \mu_r y_{r[i_T]} + \sum_{j=1}^{m} \omega_j x_{j[i_T]} \geq 0,\ T \in P \cup \lceil S \rceil, \\ & \sum_{r=1}^{l} \mu_r y_{r[i_S]} = 1, \\ & \mu_r \geq 0,\ r = 1,\dots,l,\ \omega_j \geq 0,\ j = 1,\dots,m. \end{aligned} \tag{7}$$

Next, we present a collection of theoretical results on DEA sum games that fulfil the most basic properties for partition function form games. Previously, we introduce an order relation between partitions. Recall that $\Pi(N)$ is an ordered set with the following order. Given the partitions $P, Q \in \Pi(N)$, *P precedes Q* (or *P is finer than Q*), $P \preceq Q$, if for every $S \in P$ there is $T \in Q$ such that $S \subseteq T$. In other words, the elements in Q are obtained by unions of elements in P.

A partition function form game $e \in PG(N)$ has *positive externalities* if for every $S \subseteq N$, $P, Q \in \Pi(N \setminus S)$ such that $P \preceq Q$, it satisfies $e(S; P) \leq e(S; Q)$. This means that the earning of coalition S is nondecreasing according to the ordering $\preceq$ on $\Pi(N \setminus S)$.

From the formal definition, the next result directly establishes the sense of the externalities of a DEA sum game $\mathbf{e}$.

Proposition 3.2. *Let $(N; X; Y)$ be a multiagent DEA problem. Then, the associated DEA sum game* $\mathbf{e}$ *has positive externalities.*

Proof. Let $S \subseteq N$, and $P, Q \in \Pi(N \setminus S)$ such that $P \preceq Q$. Notice that any feasible solution μ_r, $r = 1,\dots,l$, ω_j, $j = 1,\dots,m$ for Problem DP(S;P) is also feasible for Problem DP(S;Q). Then,

$$\rho^*_{[i_S],P} \geq \rho^*_{[i_S],Q}.$$

Therefore, $\mathbf{e}(S; P) \leq \mathbf{e}(S; Q)$. □

Under the DEA perspective, this property ensures that for any coalition $S \subseteq N$, its relative efficiency score increases as the elements of the partition in $N \setminus S$ progressively merge. From Proposition 3.2, it immediately follows that the value of the relative efficiency score of coalition S enlarges as DMUs outside S form larger blocks with elements of P. In particular, for every efficient DMU i, we have $\mathbf{e}(\{i\}; P) = 1$, for every $P \in \Pi(N \setminus \{i\})$.

In addition, we study the profitable character of the cooperation in DEA sum games. To do this, we consider the following property. A partition function form game e is *subadditive* (Hafalir, 2007) if, for every $S, T \subseteq N$ with $S \cap T = \emptyset$ and $P \in \Pi(N \setminus (S \cup T))$, it holds that

$$e(S \cup T; P) \leq e(S; P \cup \lceil T \rceil) + e(T; P \cup \lceil S \rceil).$$

This property ensures that the joint efficiency score of merging reduces with respect to the sum of the efficiencies of each coalition involved.

Proposition 3.3. *Let $(N; X; Y)$ be a multiagent DEA problem. The associated DEA sum game* $\mathbf{e}$ *is subadditive.*

Proof. Let $S, T \subset N$, $S \cap T = \emptyset$, and take $P \in \Pi(N \setminus (S \cup T))$. First, we check that

$$\mathbf{e}(S \cup T; P) \leq \max\left\{\mathbf{e}(S; P \cup \lceil T \rceil), \mathbf{e}(T; P \cup \lceil S \rceil)\right\}$$

to prove

$$\mathbf{e}(S \cup T; P) \leq \mathbf{e}(S; P \cup \lceil T \rceil) + \mathbf{e}(T; P \cup \lceil S \rceil).$$

If $\max\{\mathbf{e}(S; P \cup \lceil T \rceil), \mathbf{e}(T; P \cup \lceil S \rceil)\} = 1$, then the result immediately follows. Let us assume that

$$\max\left\{\mathbf{e}(S; P \cup \lceil T \rceil), \mathbf{e}(T; P \cup \lceil S \rceil)\right\} = \mathbf{e}(T; P \cup \lceil S \rceil) < 1.$$

We also assume that $\mathbf{e}(S \cup T; P) > \mathbf{e}(T; P \cup \lceil S \rceil)$. Notice that $\mathbf{e}(S \cup T; P)$ is the optimal value of the fractional programming problem

$$\begin{aligned} \max \quad & \frac{\sum_{r=1}^{l} u_r (y_{r[i_S]} + y_{r[i_T]})}{\sum_{j=1}^{m} v_j (x_{j[i_S]} + x_{j[i_T]})} \\ \text{subject to} \quad & \frac{\sum_{r=1}^{l} u_r y_{r[i_R]}}{\sum_{j=1}^{m} v_j x_{j[i_R]}} \leq 1,\ R \in P \cup \lceil S \cup T \rceil, \\ & u_r \geq 0,\ r = 1,\dots,l,\ v_j \geq 0,\ j = 1,\dots,m. \end{aligned} \tag{8}$$

Let $\hat{u}_r$, $r = 1,\dots,l$, $\hat{v}_j$, $j = 1,\dots,m$ be an optimal solution of Problem (8). Then,

$$\frac{\sum_{r=1}^{l} \hat{u}_r (y_{r[i_S]} + y_{r[i_T]})}{\sum_{j=1}^{m} \hat{v}_j (x_{j[i_S]} + x_{j[i_T]})} \leq \max\left\{ \frac{\sum_{r=1}^{l} \hat{u}_r y_{r[i_S]}}{\sum_{j=1}^{m} \hat{v}_j x_{j[i_S]}}, \frac{\sum_{r=1}^{l} \hat{u}_r y_{r[i_T]}}{\sum_{j=1}^{m} \hat{v}_j x_{j[i_T]}} \right\}.$$

We distinguish several cases.

Case 1. First, we assume that $\max\left\{ \frac{\sum_{r=1}^{l} \hat{u}_r y_{r[i_S]}}{\sum_{j=1}^{m} \hat{v}_j x_{j[i_S]}}, \frac{\sum_{r=1}^{l} \hat{u}_r y_{r[i_T]}}{\sum_{j=1}^{m} \hat{v}_j x_{j[i_T]}} \right\} \leq 1.$

Here, $\hat{u}_r$, $r = 1,\dots,l$, $\hat{v}_j$, $j = 1,\dots,m$ are feasible for coalition T and S. Then, we distinguish two subcases.

Subcase 1.1 If it satisfies that $\max\left\{ \frac{\sum_{r=1}^{l} \hat{u}_r y_{r[i_S]}}{\sum_{j=1}^{m} \hat{v}_j x_{j[i_S]}}, \frac{\sum_{r=1}^{l} \hat{u}_r y_{r[i_T]}}{\sum_{j=1}^{m} \hat{v}_j x_{j[i_T]}} \right\} = \frac{\sum_{r=1}^{l} \hat{u}_r y_{r[i_T]}}{\sum_{j=1}^{m} \hat{v}_j x_{j[i_T]}}$, then we have that

$$\mathbf{e}(T; P \cup \lceil S \rceil) \geq \frac{\sum_{r=1}^{l} \hat{u}_r y_{r[i_T]}}{\sum_{j=1}^{m} \hat{v}_j x_{j[i_T]}} \geq \mathbf{e}(S \cup T; P) > \mathbf{e}(T; P \cup \lceil S \rceil),$$

which is not possible.

Subcase 1.2 Otherwise, if it satisfies that $\max\left\{ \frac{\sum_{r=1}^{l} \hat{u}_r y_{r[i_S]}}{\sum_{j=1}^{m} \hat{v}_j x_{j[i_S]}}, \frac{\sum_{r=1}^{l} \hat{u}_r y_{r[i_T]}}{\sum_{j=1}^{m} \hat{v}_j x_{j[i_T]}} \right\} = \frac{\sum_{r=1}^{l} \hat{u}_r y_{r[i_S]}}{\sum_{j=1}^{m} \hat{v}_j x_{j[i_S]}}$, then we obtain that

$$\mathbf{e}(S \cup T; P) > \mathbf{e}(T; P \cup \lceil S \rceil) \geq \mathbf{e}(S; P \cup \lceil T \rceil) \geq \frac{\sum_{r=1}^{l} \hat{u}_r y_{r[i_S]}}{\sum_{j=1}^{m} \hat{v}_j x_{j[i_S]}} \geq \mathbf{e}(S \cup T; P).$$

However, this case is again not possible.

Case 2. Secondly, we suppose that $\max\left\{ \frac{\sum_{r=1}^{l} \hat{u}_r y_{r[i_S]}}{\sum_{j=1}^{m} \hat{v}_j x_{j[i_S]}}, \frac{\sum_{r=1}^{l} \hat{u}_r y_{r[i_T]}}{\sum_{j=1}^{m} \hat{v}_j x_{j[i_T]}} \right\} > 1.$ Without a loss of generality, we assume that

$$\max\left\{ \frac{\sum_{r=1}^{l} \hat{u}_r y_{r[i_S]}}{\sum_{j=1}^{m} \hat{v}_j x_{j[i_S]}}, \frac{\sum_{r=1}^{l} \hat{u}_r y_{r[i_T]}}{\sum_{j=1}^{m} \hat{v}_j x_{j[i_T]}} \right\} = \frac{\sum_{r=1}^{l} \hat{u}_r y_{r[i_T]}}{\sum_{j=1}^{m} \hat{v}_j x_{j[i_T]}}.$$

(Otherwise, we can naturally switch T by S for the remainder of the proof). The optimality of $\hat{u}_r$ and $\hat{v}_j$, for all $r = 1,\dots,l$ and $j = 1,\dots,m$, for obtaining $\mathbf{e}(S \cup T; P)$ ensures that

$$\frac{\sum_{r=1}^{l} \hat{u}_r (y_{r[i_S]} + y_{r[i_T]})}{\sum_{j=1}^{m} \hat{v}_j (x_{j[i_S]} + x_{j[i_T]})} \leq 1. \tag{9}$$

Notice that combining the inequality in (9) with $\frac{\sum_{r=1}^{l} \hat{u}_r y_{r[i_T]}}{\sum_{j=1}^{m} \hat{v}_j x_{j[i_T]}} \geq 1$ implies that $\sum_{r=1}^{l} \hat{u}_r y_{r[i_S]} \leq \sum_{j=1}^{m} \hat{v}_j x_{j[i_S]}$.
Take $\tilde{u}_r$, $r = 1,\dots,l$ and $\tilde{v}_j$, $j = 1,\dots,m$, which is an optimal solution for Problem (10) that gives $e(T; P \cup \lceil S \rceil)$,

$$\begin{aligned} \max \quad & \frac{\sum_{r=1}^{l} u_r y_{r[i_T]}}{\sum_{j=1}^{m} v_j x_{j[i_T]}} \\ \text{subject to} \quad & \frac{\sum_{r=1}^{l} u_r y_{r[i_R]}}{\sum_{j=1}^{m} v_j x_{j[i_R]}} \leq 1,\ R \in P \cup \lceil S \rceil \cup \lceil T \rceil, \\ & u_r \geq 0,\ r = 1,\dots,l,\ v_j \geq 0,\ j = 1,\dots,m. \end{aligned} \tag{10}$$

In addition, define $\bar{u}_r = \frac{\hat{u}_r}{K}$, with $K = \frac{\sum_{r=1}^{l} \hat{u}_r y_{r[i_T]}}{\sum_{j=1}^{m} \hat{v}_j x_{j[i_T]}}$, for all $r = 1, \dots, l$ and $\bar{v}_j = \hat{v}_j$, for all $j = 1, \dots, m$. We check that these elements are also feasible for Problem (10). It is clear that $\bar{u}_r \geq 0$, $r = 1, \dots, l$ and $\bar{v}_j \geq 0$, $j = 1, \dots, m$.
Take $R \in P$. Then,

$$\frac{\sum_{r=1}^{l} \bar{u}_r y_{r[i_R]}}{\sum_{j=1}^{m} \bar{v}_j x_{j[i_R]}} = \frac{\sum_{r=1}^{l} \frac{\hat{u}_r}{K} y_{r[i_R]}}{\sum_{j=1}^{m} \hat{v}_j x_{j[i_R]}} = \frac{1}{K} \frac{\sum_{r=1}^{l} \hat{u}_r y_{r[i_R]}}{\sum_{j=1}^{m} \hat{v}_j x_{j[i_R]}} \leq 1$$

because $K = \frac{\sum_{r=1}^{l} \hat{u}_r y_{r[i_T]}}{\sum_{j=1}^{m} \hat{v}_j x_{j[i_T]}} > 1$ and the feasibility condition of $\hat{u}_r$, $r = 1, \dots, l$ and $\hat{v}_j$, $j = 1, \dots, m$ for any $R \in P$.
Take $R = S$. Then,

$$\frac{\sum_{r=1}^{l} \bar{u}_r y_{r[i_S]}}{\sum_{j=1}^{m} \bar{v}_j x_{j[i_S]}} = \frac{\sum_{r=1}^{l} \frac{\hat{u}_r}{K} y_{r[i_S]}}{\sum_{j=1}^{m} \hat{v}_j x_{j[i_S]}} = \frac{1}{K} \frac{\sum_{r=1}^{l} \hat{u}_r y_{r[i_S]}}{\sum_{j=1}^{m} \hat{v}_j x_{j[i_S]}} \leq 1$$

because $K = \frac{\sum_{r=1}^{l} \hat{u}_r y_{r[i_T]}}{\sum_{j=1}^{m} \hat{v}_j x_{j[i_T]}} > 1$ and the condition $\sum_{r=1}^{l} \hat{u}_r y_{r[i_S]} \leq \sum_{j=1}^{m} \hat{v}_j x_{j[i_S]}$.
Finally, take $R = T$. Then,

$$\frac{\sum_{r=1}^{l} \bar{u}_r y_{r[i_T]}}{\sum_{j=1}^{m} \bar{v}_j x_{j[i_T]}} = \frac{\sum_{r=1}^{l} \frac{\hat{u}_r}{K} y_{r[i_T]}}{\sum_{j=1}^{m} \hat{v}_j x_{j[i_T]}} = \frac{1}{K} \frac{\sum_{r=1}^{l} \hat{u}_r y_{r[i_T]}}{\sum_{j=1}^{m} \hat{v}_j x_{j[i_T]}} = 1$$

because $K = \frac{\sum_{r=1}^{l} \hat{u}_r y_{r[i_T]}}{\sum_{j=1}^{m} \hat{v}_j x_{j[i_T]}}$. In addition, it implies that

$$\frac{\sum_{r=1}^{l} \bar{u}_r y_{r[i_T]}}{\sum_{j=1}^{m} \bar{v}_j x_{j[i_T]}} = 1 > \mathbf{e}(T; P \cup \lceil S \rceil) = \frac{\sum_{r=1}^{l} \tilde{u}_r y_{r[i_T]}}{\sum_{j=1}^{m} \tilde{v}_j x_{j[i_T]}}.$$

This fact contradicts the optimality of $\tilde{u}_r$, $r = 1, \dots, l$ and $\tilde{v}_j$, $j = 1, \dots, m$, and $\mathbf{e}(T; P \cup \lceil S \rceil)$.

Then, $\mathbf{e}(S \cup T; P) \leq \max\{\mathbf{e}(S; P \cup \lceil T \rceil), \mathbf{e}(T; P \cup \lceil S \rceil)\} \leq \mathbf{e}(S; P \cup \lceil T \rceil) + \mathbf{e}(T; P \cup \lceil S \rceil)$, thus concluding the proof. □

Games with externalities deal with coalitions and allocations, and they consider groups of agents willing to allocate the joint benefits derived from their cooperation. In the remainder of this paper, we will use allocations for games with externalities to rank DMUs. More specifically, we will consider proposals that make use of games with transferable utility for this purpose. Recall that a *game with transferable utility or TU game* is a pair (N, v), where N is a finite set of agents and v is a map from 2^N to $\mathbb{R}$ that satisfies that $v(\emptyset) = 0$ (cf. González-Díaz et al., 2010). The class of TU games with a set of agents N is denoted by G^N. Given $e \in PG(N)$, we can associate some intuitive TU games. For instance, considering an optimistic perspective, we can define the TU game $(N, e_{\max}) \in G^N$ for every $S \subseteq N$, as

$$e_{\max}(S) = \max_{P \in \Pi(N \setminus S)} e(S; P)$$

From a pessimistic point of view, we can define the TU game $(N, e_{\min}) \in G^N$. It is formally given for every $S \subseteq N$ by

$$e_{\min}(S) = \min_{P \in \Pi(N \setminus S)} e(S; P).$$

By their definition, $e_{\max}$ and $e_{\min}$ assign to each $S \subseteq N$ the maximum and the minimum value that the members of S can obtain by their cooperation, respectively, among all possible structures of $N \setminus S$.

Alternatively, we can consider averaging procedures. Albizuri et al. (2005) proposed assigning to every $S \subseteq N$ the TU game $(N, \bar{e}) \in G^N$ given by

$$\bar{e}(S) = \frac{1}{|\Pi(N \setminus S)|} \sum_{P \in \Pi(N \setminus S)} e(S; P). \quad (11)$$

Here, each coalition $S \subseteq N$ obtains the expected value of the cooperation of the members of S in e, that is, the average over the whole set of the embedded coalitions $(S; P)$ with $P \in \Pi(N \setminus S)$ when they are equally likely.

Finally, we mention the approach of Hu and Yang (2010), which consider the TU game $(N, \bar{\bar{e}}) \in G^N$ given by

$$\bar{\bar{e}}(S) = \frac{1}{|\Pi(N)|} \sum_{P \in \Pi(N)} e(S; P_{-S}), \text{ for each } S \subseteq N. \quad (12)$$

Thus, for each $S \subseteq N$, $\bar{\bar{e}}(S)$ is the expected worth of S in e over the set of embedded coalitions induced by any partition in $\Pi(N)$, assuming that all $P \in \Pi(N)$ are equally likely.

In our case, we have explicit expressions for the optimistic and the pessimistic TU games $\mathbf{e}_{\max}$ and $\mathbf{e}_{\min}$. From Proposition 3.2, it readily follows for any DEA sum game $\mathbf{e}$ that

$$\mathbf{e}_{\max}(S) = \mathbf{e}(S; \lceil N \setminus S \rceil) \text{ and } \mathbf{e}_{\min}(S) = \mathbf{e}(S; \lfloor N \setminus S \rfloor).$$

This implies that the maximum worth of the cooperation of S is reached when the outsiders act as a whole block, while the minimum is reached when the partition of $N \setminus S$ is formed by the individual agents.

We illustrate the computation of $(N, \mathbf{e}_{max})$, $(N, \mathbf{e}_{min})$, $(N, \bar{\mathbf{e}})$ and $(N, \bar{\bar{\mathbf{e}}})$ in our setting. It also allows us to discuss some of their properties. A more exhaustive analysis of properties of the resulting TU games could be carried out, but this is beyond the scope of ranking DMUs.

Example 3.4. Let $(N; X; Y)$ be a multiagent DEA problem, $N = \{1, 2, 3, 4\}$. We consider the case with a single input and a single output given by the data set of the inputs and the outputs of the DMUs described below.

$$X_1 = (30), \ X_2 = (1), \ X_3 = (3), \text{ and } X_4 = (11).$$

$$Y_1 = (17), \ Y_2 = (27), \ Y_3 = (22), \text{ and } Y_4 = (9).$$

Thus, the associated DEA sum game $\mathbf{e}$ is given by

$\mathbf{e}(N; \emptyset) = 1$, $\mathbf{e}(\{1,2,4\}; \{3\}) = 0.153$, $\mathbf{e}(\{3\}; \{1,2,4\}) = 1$,
$\mathbf{e}(\{1,2,3\}; \{4\}) = 1$, $\mathbf{e}(\{4\}; \{1,2,3\}) = 0.464$, $\mathbf{e}(\{1,3,4\}; \{2\}) = 0.035$,
$\mathbf{e}(\{2\}; \{1,3,4\}) = 1$, $\mathbf{e}(\{2,3,4\}; \{1\}) = 1$, $\mathbf{e}(\{1\}; \{2,3,4\}) = 0.0948$,
$\mathbf{e}(\{1,4\}; \{2,3\}) = 0.040$, $\mathbf{e}(\{2,3\}; \{1,4\}) = 1$, $\mathbf{e}(\{1,2\}; \{3,4\}) = 0.554$,
$\mathbf{e}(\{3,4\}; \{1,2\}) = 1$, $\mathbf{e}(\{1,3\}; \{2,4\}) = 0.333$, $\mathbf{e}(\{2,4\}; \{1,3\}) = 1$,
$\mathbf{e}(\{1,4\}; \lfloor 2,3 \rfloor) = 0.018$, $\mathbf{e}(\{2\}; \{\{1,4\},\{3\}\}) = 1$, $\mathbf{e}(\{3\}; \{\{1,4\},\{2\}\}) = 0.272$,
$\mathbf{e}(\{1,2\}; \lfloor 3,4 \rfloor) = 0.167$, $\mathbf{e}(\{3\}; \{\{1,2\},\{4\}\}) = 1$, $\mathbf{e}(\{4\}; \{\{1,2\},\{3\}\}) = 0.112$,
$\mathbf{e}(\{1,3\}; \lfloor 2,4 \rfloor) = 0.037$, $\mathbf{e}(\{2\}; \{\{1,3\},\{4\}\}) = 1$, $\mathbf{e}(\{4\}; \{\{1,3\},\{2\}\}) = 0.030$,
$\mathbf{e}(\{2,4\}; \lfloor 1,3 \rfloor) = 0.409$, $\mathbf{e}(\{1\}; \{\{2,4\},\{3\}\}) = 0.05$, $\mathbf{e}(\{3\}; \{\{2,4\},\{1\}\}) = 1$,
$\mathbf{e}(\{2,3\}; \lfloor 1,4 \rfloor) = 1$, $\mathbf{e}(\{1\}; \{\{2,3\},\{4\}\}) = 0.030$, $\mathbf{e}(\{4\}; \{\{2,3\},\{1\}\}) = 0.067$
$\mathbf{e}(\{3,4\}; \lfloor 1,2 \rfloor) = 0.082$, $\mathbf{e}(\{1\}; \{\{3,4\},\{2\}\}) = 0.014$, $\mathbf{e}(\{2\}; \{\{3,4\},\{1\}\}) = 1$,
$\mathbf{e}(\{1\}; \lfloor 2,3,4 \rfloor) = 0.014$, $\mathbf{e}(\{2\}; \lfloor 1,3,4 \rfloor) = 1$, $\mathbf{e}(\{3\}; \lfloor 1,2,4 \rfloor) = 0.272$,
$\mathbf{e}(\{4\}; \lfloor 1,2,3 \rfloor) = 0.030$.

Table 1 depicts the TU games $(N, \bar{\mathbf{e}})$, $(N, \mathbf{e}_{min})$, $(N, \mathbf{e}_{max})$ and $(N, \bar{\bar{\mathbf{e}}})$ associated with the DEA sum game $\mathbf{e}$.

Table 1
TU games $(N, \bar{\mathbf{e}})$, $(N, \mathbf{e}_{min})$, $(N, \mathbf{e}_{max})$, and $(N, \bar{\bar{\mathbf{e}}})$.

S	$\bar{\mathbf{e}}(S)$	$\mathbf{e}_{min}(S)$	$\mathbf{e}_{max}(S)$	$\bar{\bar{\mathbf{e}}}(S)$	S	$\bar{\mathbf{e}}(S)$	$\mathbf{e}_{min}(S)$	$\mathbf{e}_{max}(S)$	$\bar{\bar{\mathbf{e}}}(S)$
$\emptyset$	0.000	0.000	0.000	0.000	$\{2,3\}$	1.000	1.000	1.000	1.000
$\{1\}$	0.062	0.021	0.147	0.054	$\{2,4\}$	0.705	0.409	1.000	0.606
$\{2\}$	1.000	1.000	1.000	1.000	$\{3,4\}$	0.541	0.082	1.000	0.388
$\{3\}$	0.709	0.272	1.000	0.660	$\{1,2,3\}$	1.000	1.000	1.000	1.000
$\{4\}$	0.132	0.030	0.422	0.106	$\{1,2,4\}$	0.172	0.172	0.172	0.172
$\{1,2\}$	0.417	0.194	0.641	0.343	$\{1,3,4\}$	0.040	0.040	0.040	0.040
$\{1,3\}$	0.219	0.044	0.394	0.161	$\{2,3,4\}$	1.000	1.000	1.000	1.000
$\{1,4\}$	0.038	0.024	0.052	0.033	N	1.000	1.000	1.000	1.000

It is easy to verify that $\mathbf{e}_{max}(S \cup T) \leq \mathbf{e}_{max}(S) + \mathbf{e}_{max}(T)$ for every S, $T \subseteq N$ with $S \cap T = \emptyset$. However, this property does not hold for $\mathbf{e}_{min}$, $\bar{\mathbf{e}}$, nor $\bar{\bar{\mathbf{e}}}$. For instance, take $S = \{1,2\}$ and $T = \{3,4\}$. ◁

4. Ranking DMUs under externalities

In this section, we address the task of ranking DMUs in multiagent DEA problems under the idea of cooperation described above.

In the literature, the relative efficiency score of DMUs is frequently used as a criterion. For this purpose, some tools taken from cooperative game theory, such as the values or solutions for TU games, are used. They assign a vector in $\mathbb{R}^N$ to every TU game $(N, v) \in G^N$ and they are usually based on the notion of *marginal contribution* of an agent to those coalitions that do not contain it. That is, fixed $i \in N$ and $(N, v) \in G^N$, agent i's marginal contribution to $T \subseteq N \setminus \{i\}$ is given as follows:

$$v(T \cup \{i\}) - v(T). \tag{13}$$

One of the most important values is the *Shapley value* (Shapley, 1953), which is defined, for every $i \in N$ and every TU game $(N, v) \in G^N$, as follows:

$$Sh_i(N, v) = \sum_{T \subseteq N \setminus \{i\}} \frac{|T|!\,(|N| - |T| - 1)!}{|N|!}(v(T \cup \{i\}) - v(T)). \tag{14}$$

Under externalities, natural extensions of this value are observed. In DEA settings, these solutions provide a measure of the ability of DMUs to modify the relative efficiency score of merging under cooperation.

Let $e \in PG(N)$ be a game with externalities. In what follows, we list the values proposed for games in partition function form that we will use to rank DMUs. Each of them, even those defined as an allocation that satisfy some properties, corresponds to the Shapley value of one of the TU games described at the end of Section 3.

- First, we refer to Albizuri et al. (2005), who propose as a solution for e the Shapley value for the TU game $(N, \bar{e})$, that is, $Sh(N, \bar{e})$.
- The solution proposed by de Clippel and Serrano (2008) and Pham Do and Norde (2007) for e is the Shapley value for the associated TU game (N, e_{min}), that is, $Sh(N, e_{min})$. It is known as the externality-free Shapley value.
- McQuillin (2009) proposes an allocation represented by the Shapley value for the TU game (N, e_{max}), that is, $Sh(N, e_{max})$.
- Finally, we consider the approach of Hu and Yang (2010) who consider the Shapley value for $(N, \bar{\bar{e}})$.

Below, Example 4.1 illustrates their performance on DEA sum games.

Example 4.1. Let $(N; X; Y)$ be the multiagent DEA problem, with $N = \{1, 2, 3, 4\}$ and two inputs and a single output. The associated data set for the DMUs in N is given as follows:

$$X_1 = \begin{pmatrix} 2 \\ 9 \end{pmatrix},\; X_2 = \begin{pmatrix} 10 \\ 8 \end{pmatrix},\; X_3 = \begin{pmatrix} 6 \\ 7 \end{pmatrix},\; \text{and } X_4 = \begin{pmatrix} 4 \\ 5 \end{pmatrix};$$

and

$$Y_1 = (4),\; Y_2 = (10),\; Y_3 = (9),\; \text{and } Y_4 = (6).$$

Table 2 depicts the characteristic function of the above-mentioned TU games $(N, \bar{\mathbf{e}})$, $(N, \mathbf{e}_{min})$, $(N, \mathbf{e}_{max})$ and $(N, \bar{\bar{\mathbf{e}}})$, which are associated with the corresponding game $\mathbf{e}$ with externalities.

Table 2
TU games $(N, \bar{\mathbf{e}})$, $(N, \mathbf{e}_{min})$, $(N, \mathbf{e}_{max})$, and $(N, \bar{\bar{\mathbf{e}}})$.

S	$\bar{\mathbf{e}}(S)$	$\mathbf{e}_{min}(S)$	$\mathbf{e}_{max}(S)$	$\bar{\bar{\mathbf{e}}}(S)$	S	$\bar{\mathbf{e}}(S)$	$\mathbf{e}_{min}(S)$	$\mathbf{e}_{max}(S)$	$\bar{\bar{\mathbf{e}}}(S)$
$\emptyset$	0.000	0.000	0.000	0.000	$\{2,3\}$	1.000	1.000	1.000	1.000
$\{1\}$	1.000	1.000	1.000	1.000	$\{2,4\}$	0.979	0.957	1.000	0.972
$\{2\}$	0.989	0.972	1.000	0.987	$\{3,4\}$	1.000	1.000	1.000	1.000
$\{3\}$	1.000	1.000	1.000	1.000	$\{1,2,3\}$	0.852	0.852	0.852	0.852
$\{4\}$	0.998	0.992	1.000	0.998	$\{1,2,4\}$	0.833	0.833	0.833	0.833
$\{1,2\}$	0.778	0.778	0.778	0.778	$\{1,3,4\}$	1.000	1.000	1.000	1.000
$\{1,3\}$	1.000	1.000	1.000	1.000	$\{2,3,4\}$	1.000	1.000	1.000	1.000
$\{1,4\}$	1.000	1.000	1.000	1.000	N	1.000	1.000	1.000	1.000

The row in Table 3 named "Eff." includes the individual relative efficiency score, which is given by $\mathbf{e}(\{i\}; \lfloor N \setminus \{i\} \rfloor)$ for each $i \in N$, as well as the corresponding ranking. Below, we also include the ranking based on the allocations proposed in Albizuri et al. (2005), which is denoted by (A), that proposed in de Clippel and Serrano (2008), which is denoted by (CS), that proposed in McQuillin (2009), which is denoted by (MQ), and that proposed by Hu and Yang (2010), which is denoted by (HY). In the remainder of this section, for DMUs that are indistinguishable in a ranking (i.e., showing the same relative efficiency score), we will assign the artificial position given by the average of their positions. That is, positions of DMUs 1 and 3 that occupy the first two positions in the ranking of individual efficiencies, are denoted in the ranking by 1.5.

Table 3
Ranking of DMUs.

	Numerical results				Position of DMUs			
	DMU 1	DMU 2	DMU 3	DMU 4	DMU 1	DMU 2	DMU 3	DMU 4
Eff.	1.000	0.972	1.000	0.992	1.5	4	1.5	3
(A)	0.208	0.201	0.301	0.290	3	4	1	2
(CS)	0.212	0.195	0.304	0.288	3	4	1	2
(MQ)	0.205	0.205	0.298	0.292	3	4	1	2
(HY)	0.209	0.200	0.301	0.290	3	4	1	2

In this example, DMU 4 has a higher rank than DMU 1, despite having a lower relative efficiency score than DMU 1. ◁

All four rankings coincide in the previous example, which is not usual. The following example, with 12 DMUS, illustrates this fact.

Example 4.2. We revisit Example 1 taken from Kritikos (2017). The data set of the inputs and the outputs is shown in Table 4. The individual efficiencies under the DEA method are in the second column of Table 5.

Table 4
Input and output data in Example 1 of Kritikos (2017).

DMU	Input 1	Input 2	Output 1	Output 2	Output 3	Output 4
1	17.02	5.0	42	45.3	14.2	30.1
2	16.46	4.5	39	40.1	13.0	29.8
3	11.76	6.0	26	39.6	13.8	24.5
4	10.52	4.0	22	36.0	11.3	25.0
5	9.50	3.8	21	34.2	12.0	20.4
6	4.79	5.4	10	20.1	5.0	16.5
7	6.21	6.2	14	26.5	7.0	19.7
8	11.12	6.0	25	35.9	9.0	24.7
9	3.67	8.0	4	17.4	0.1	18.1
10	8.93	7.0	16	34.3	6.5	20.6
11	17.74	7.1	43	45.6	14.0	31.1
12	14.85	6.2	27	38.7	13.8	25.4

Table 5
Ranking of DMUs.

DMU	Efficiency	Rank	(A)	Rank	(CS)	Rank	(MQ)	Rank
1	1.000	4	0.08676	5	0.09377	4	0.08323	7
2	1.000	4	0.08462	6	0.08720	6	0.08313	8
3	0.982	9	0.08453	7	0.08254	7	0.08348	5
4	1.000	4	0.08979	3	0.09656	3	0.08392	2
5	1.000	4	0.09049	2	0.10007	2	0.08379	4
6	1.000	4	0.08862	4	0.09295	5	0.08381	3
7	1.000	4	0.09183	1	0.10361	1	0.08411	1
8	0.961	10	0.08317	8	0.08028	9	0.08348	6
9	1.000	4	0.07685	11	0.06913	10	0.08267	11
10	0.954	11	0.07706	10	0.06866	11	0.08292	9
11	0.983	8	0.08165	9	0.08167	8	0.08282	10
12	0.801	12	0.06464	12	0.04357	12	0.08263	12

From a computational perspective, the solutions of Albizuri et al. (2005), de Clippel and Serrano (2008), and McQuillin (2009), denoted by (A), (CS) and (MQ), respectively, can be exactly obtained in a reasonable time. Consequently, we are able to rank DMUs from the ordering of their components, and they are all depicted in Table 5.

Below, we provide comments on the rankings. The proposals of (A) and (CS) assign a relative efficiency score equal to 1 to the first six positions of the DMU rankings, and the less efficient DMU (DMU 12) occupies the last position. Under the proposal of (MQ), not all DMUs with a relative efficiency score equal to 1 occupy the top positions (see, for instance, the cases of DMU 3 and DMU 8 in positions 5 and 6), although DMU 12 is again in the last position.

Table 6
Spearman's correlation matrix for the rankings of DMUs.

	(A)	(CS)	(MQ)
(A)	1.00000	0.97902	0.91608
(CS)	0.97902	1.00000	0.84615
(MQ)	0.91608	0.84615	1.00000

Using the Spearman's correlation coefficients depicted in Table 6, we observe that the pairs of more similar rankings are (A) and (CS), (A) and (MQ), and (CS) and (MQ), respectively. ◃

The previous example reflects that the proposal of Hu and Yang (2010) cannot be obtained exactly in a reasonable calculation time, even in those cases with a relatively small number of agents, such as in Example 4.2, since the total amount of partitions in $\Pi(N)$ is equal to 4 213 597 for every coalition. These drawbacks may also arise for the case of Albizuri et al. (2005) for coalitions of small sizes. For instance, only having 13 agents instead of 12 would imply that the number of partitions to be evaluated for those coalitions of the form $\{\{i\} : i \in N\}$ is equal to 4 213 597, which makes difficult to obtain the exact value of $\bar{e}(\{i\})$ for every $i \in N$. Nevertheless, we can exactly obtain this value in Example 4.2. Table 7 shows $|\Pi(N \setminus S)|$, for every $S \subseteq N$.

Table 7
Number of partitions for several coalition sizes for 12 agents.

$\|S\|$	$\|N \setminus S\|$	$\|\Pi(N \setminus S)\|$	$\|S\|$	$\|N \setminus S\|$	$\|\Pi(N \setminus S)\|$	$\|S\|$	$\|N \setminus S\|$	$\|\Pi(N \setminus S)\|$
1	11	678 570	5	7	877	9	3	5
2	10	115 975	6	6	203	10	2	2
3	9	21 147	7	5	52	11	1	1
4	8	4140	8	4	15	12	0	1

5. Estimating rankings of DMUs under externalities

In this section, we address those computational problems that arise when calculating some of the above-mentioned solutions for games with externalities. Along with the usual problems observed for calculating the Shapley value with a large number of agents (see Castro et al. (2009) or Fernández-García and Puerto-Albandoz (2006)), new difficulties arise only in determining the characteristic functions of some considered games, as we indicate in Section 4 for the TU games $\bar{e}$ and $\bar{\bar{e}}$. Below, we describe two procedures based on sampling techniques that reduce this complexity for any game with externalities e involving a large amount of agents.

The numerical results included in the following sections have been computed using the statistical software R (R Core Team, 2021) on a personal computer with an Intel(R) Core(TM) i9-9980HK, 32 GB of memory and a single 2.40 GHz CPU processor. Specifically, the *lpSolveAPI* package in R software, which is used to solve mathematical linear programming problems, and the *numbers* package and *partitions* package, which are used for managing with partitions, are required.

5.1. Estimating the ranking of Hu and Yang

Here, we describe a sampling algorithm for ranking DMUs according to the Shapley value for the TU game $(N, \bar{\bar{e}})$ of Hu and Yang (2010). In particular, this proposal is mainly focused on the estimation of its characteristic function.

Let e be a partition function form game. Under this approach, the worth of the cooperation of DMUs in a given coalition S is the expected value of the game e with externalities on the set of partitions of N (see Expression (12)). Thus, with fixed $S \subseteq N$, the problem of estimating $\bar{\bar{e}}(S)$ is a very common task in Statistics. For this purpose, simple random sampling with replacement (Cochran, 2007) on the set of partitions of N is useful. Therefore, for a given sample of partitions, the estimation of $\bar{\bar{e}}(S)$, $\widehat{\widehat{e_S}}$, is the average of the worth of the game in partition function form over the partitions for $N \setminus S$ induced by the sampled units. Recall that, if we take $P \in \Pi(N)$, P induces the partition for $N \setminus S$ given by $P_{-S} = \{T \setminus S : T \in P\}$.

The proposed sampling procedure is described as follows.

1. The parameter under study is $\bar{\bar{e}}(S)$ for a fixed $S \subseteq N$.
2. The population of the sampling procedure is the set of partitions $\Pi(N)$.
3. The characteristic under study for each sampling unit $P \in \Pi(N)$ is the worth of coalition S in e when agents in $N \setminus S$ are joined according to the partition induced by P for $N \setminus S$, which is denoted by P_{-S}. Thus, the sampling unit is P_{-S}.
4. The sampling procedure takes each partition $P \in \Pi(N)$ with the same probability. The process chooses at random a partition P^k for all $k \in \{1, \dots, |\Pi(N)|\}$. Thus, with replacement, we obtain a sample $S_P = \{P_1, \dots, P_\ell\}$ of size ℓ.
5. The estimation of $\bar{\bar{e}}(S)$ with $S \subseteq N$ is the mean of $e(S; P_{-S})$ over the sample of partitions P in S_P. That is, $\widehat{\widehat{e_S}} = \frac{1}{\ell} \sum_{P \in S_P} e(S; P_{-S})$ for each $S \subseteq N$, with ℓ denoting the sample size such that $1 < \ell \leq |\Pi(N)|$.

The pseudocode to estimate $\bar{\bar{e}}(S)$ for a given $S \subseteq N$ is depicted in Procedure 5.1.

Procedure 5.1. Take $e \in PG(N)$. Fix $S \subseteq N$.
Set ℓ. Do $j = 0$ and $\widehat{\widehat{e_S}} = 0$.
while $j < \ell$ **do**
Do $j = j + 1$ and take $P \in \Pi(N)$ with replacement.
Do $\widehat{\widehat{e_S}} = \widehat{\widehat{e_S}} + e(S; P_{-S})$.
end while
Finally, $\widehat{\widehat{e_S}} = \frac{\widehat{\widehat{e_S}}}{\ell}$.

We highlight that $\widehat{\widehat{e_S}}$ is an unbiased and consistent estimator for $\bar{\bar{e}}(S)$. Following Saavedra-Nieves et al. (2018), the task of establishing a probabilistic bound of error in estimating $\bar{\bar{e}}(S)$, with $S \subseteq N$, can be addressed. In a natural manner, the use of the estimation of $(N, \bar{\bar{e}})$, given by $\{\widehat{\widehat{e_S}}\}_{S \subseteq N}$, in (14) provides an estimator of the i component of the Shapley value for $(N, \bar{\bar{e}})$. We refer the reader to Appendix A.1 in the Online Resource Section for a detailed collection of statistical results focused on these topics.

We evaluate how this sampling proposal performs in ranking DMUs in Example 4.2.

Example 5.2. We rank the DMUs in Example 4.2 under the approach in Hu and Yang (2010) with sampling. The ratio between the amount of sampled partitions and the population size, or the sampling fraction, is denoted by f. Here, f is constant because the sampling population does not change.

Table 8 depicts the Shapley value for the estimated TU game of Hu and Yang (2010) in columns 4, 6, and 8, with f equal to 10^{-4}, $5 \cdot 10^{-4}$, and 10^{-3}. In practice, these values correspond to sample sizes ℓ (after rounding to the upper integer) of 422, 2107, and 4214 partitions of N. These amounts ensure that for each $S \subseteq N$, the bounds on the error in estimating $\bar{\bar{e}}(S)$ are equal to 0.06611, 0.02959, and 0.02092, by using the inequality (A.1.1) in Appendix A.1 in the Online Resource Section, with $\alpha = 0.05$. In addition, for each $i \in N$, the error in the Shapley value estimation ($\sqrt{\mathrm{Var}(\widehat{\widehat{Sh_i}})}$) is also bounded by 0.00994, 0.00445, and 0.00314, respectively.

The rankings obtained place those DMUs with relative efficiency score equal to 1 in the top positions except for DMU 9, which is ranked tenth among those DMUs with efficiency score less than 1. However, these ranking still respect the position of the less efficient DMU, which is the twelfth one. ◃

Table 8
Ranking of DMUs using the individual efficiencies and the estimated Shapley value for the TU game of Hu and Yang (2010).

DMU	Efficiency		Estimated rankings					
	Eff.	Rank	$f = 10^{-4}$	Rank	$f = 5 \cdot 10^{-4}$	Rank	$f = 10^{-3}$	Rank
1	1.000	4	0.093322	3	0.093292	3	0.093288	3
2	1.000	4	0.086539	6	0.086542	6	0.086577	6
3	0.982	9	0.084275	7	0.084251	7	0.084237	7
4	1.000	4	0.093236	4	0.093194	4	0.093204	4
5	1.000	4	0.095728	1	0.095675	1	0.095698	1
6	1.000	4	0.088152	5	0.088117	5	0.088133	5
7	1.000	4	0.093964	2	0.093980	2	0.093963	2
8	0.961	10	0.081901	9	0.081844	9	0.081840	9
9	1.000	4	0.075715	10	0.075829	10	0.075817	10
10	0.954	11	0.072909	11	0.073060	11	0.073085	11
11	0.983	4	0.083424	8	0.083413	8	0.083417	8
12	0.801	12	0.050835	12	0.050803	12	0.050739	12

5.2. Estimating the ranking of Albizuri et al.

Consistent with the previous section, we now analyse the problem of ranking DMUs according to the Shapley value for $(N, \overline{e})$, i.e. the TU game of Albizuri et al. (2005). For this purpose, we also consider the estimation of its characteristic function.

Take e as a partition function form game. For $S \subseteq N$, the worth of the cooperation of DMUs in S is the expected value of the game e with externalities, although it now applies over the set of partitions of $N \setminus S$. Roughly speaking, the estimation of $\overline{e}(S)$, $\widehat{e_S}$, corresponds to the sample mean of $e(S; P)$, in which P is an element of a sample of partitions in $\Pi(N \setminus S)$ obtained under simple random sampling with replacement (cf. Cochran, 2007).

The steps of our sampling proposal are as follows:

1. The parameter under study is $\overline{e}(S)$ for a fixed $S \subseteq N$.
2. The population of the sampling procedure is the set of partitions $\Pi(N \setminus S)$.
3. The characteristic under study for each sampling unit $P \in \Pi(N \setminus S)$ is the worth of coalition S in e, when agents in $N \setminus S$ are joined according to P. Thus, the sampling unit will be P.
4. The sampling procedure takes each partition $P \in \Pi(N \setminus S)$ with the same probability. A partition P^k is chosen at random for all $k \in \{1, \dots, |\Pi(N \setminus S)|\}$. Thus, we obtain with replacement the sample $S_{\mathcal{P}} = \{P_1, \dots, P_{\ell_S}\}$, with size ℓ_S.
5. The estimation of $\overline{e}(S)$, with $S \subseteq N$, corresponds to the mean of $e(S; P)$ over the sample $S_{\mathcal{P}}$. That is, $\widehat{e_S} = \frac{1}{\ell_S} \sum_{P \in S_{\mathcal{P}}} e(S; P)$ for each $S \subseteq N$, being ℓ_S the sample size with $1 < \ell_s \leq |\Pi(N \setminus S)|$.

Procedure 5.3 shows the pseudocode for estimating $\overline{e}(S)$ for a fixed $S \subseteq N$.

Procedure 5.3. Take $e \in PG(N)$. Fix $S \subseteq N$.
Set ℓ_S. Do $j = 0$ and $\widehat{e_S} = 0$.
while $j < \ell_S$ **do**
Do $j = j + 1$ and take $P \in \Pi(N \setminus S)$ with replacement.
Do $\widehat{e_S} = \widehat{e_S} + e(S; P)$.
end while
Finally, $\widehat{e_S} = \frac{\widehat{e_S}}{\ell_S}$.

For a certain coalition $S \subseteq N$, the estimator $\widehat{e_S}$ is also unbiased and consistent. Under this perspective, for every $S \subseteq N$, the task of bounding the error in estimating $\overline{e}(S)$ can be also addressed. Now, sample sizes need to be different per coalition to ensure an equitable computational effort in coalitions because the number of involved partitions changes with the cardinal of coalitions, unlike the case of Hu and Yang (2010). An estimator of the Shapley value for the TU game $(N, \overline{e})$ of Albizuri et al. (2005) is given for every $i \in N$ by the Shapley value for the estimation of $(N, \overline{e})$, which is given by the collection $\{\widehat{e_S}\}_{S \subseteq N}$. Appendix A.2 in the Online Resource Section formally details all the evidence supporting these conclusions.

Finally, we estimate the rankings of DMUs in Example 4.2 using this approach.

Example 5.4. We reconsider Example 4.2 under the approach of Albizuri et al. (2005) with sampling. We take sampling fraction $f_S = 0.1$ for each $S \subseteq N$ such that $1 \leq |S| \leq 5$. Table 9 depicts the theoretical bounds of the error in estimating $\overline{e}(S)$, with $\alpha = 0.05$ (see Inequality A.2.1 in the Online Resource Section).

Table 9
Theoretical errors (ϵ) for estimating $(N, \overline{e})$.

$\|S\|$	1	2	3	4	5
Th. error	$5.214 \cdot 10^{-3}$	$1.261 \cdot 10^{-2}$	$2.953 \cdot 10^{-2}$	$6.675 \cdot 10^{-2}$	$1.448 \cdot 10^{-1}$

For each $S \subseteq N$, with $|S| > 5$, we are completely exhaustive in evaluating the embedded coalitions $(S; P)$, with $P \in \Pi(N \setminus S)$ (see Table 7).

Table 10
Ranking of DMUs using the individual efficiencies, the exact Shapley value and its estimation for the TU game of Albizuri et al. (2005).

DMU	Efficiency		Albizuri et al.		Estimated ranking		
	Eff.	Rank	(A)	Rank	$f_S = 10^{-1}$	Rank	Abs. error
1	1.000	4	0.08676	5	0.09067	4	0.00391
2	1.000	4	0.08462	6	0.08617	6	0.00155
3	0.982	9	0.08453	7	0.08513	7	0.00060
4	1.000	4	0.08979	3	0.09160	2	0.00181
5	1.000	4	0.09049	2	0.09268	1	0.00219
6	1.000	4	0.08862	4	0.08759	5	0.00103
7	1.000	4	0.09183	1	0.09131	3	0.00052
8	0.961	10	0.08317	8	0.08275	9	0.00042
9	1.000	4	0.07685	11	0.07801	10	0.00116
10	0.954	11	0.07706	10	0.07563	11	0.00143
11	0.983	4	0.08165	9	0.08314	8	0.00149
12	0.801	12	0.06464	12	0.05532	12	0.00932

Table 10 shows the exact Shapley value as well as its estimation. The average absolute error of the estimations is equal to 0.00212, clearly less than $\sqrt{\text{Var}(\widehat{Sh_i})}$ which in this particular example can be bounded by 0.01044 (see the Online Resource Section for details). The positions given by the estimations (column 7) vary by at most one or two positions with respect to the exact ranking (column 3). The Spearman's rank correlation coefficient between the exact and the estimated ranking is 0.95804. ◁

6. An application: the hotel industry in Spain

In this section, we apply the proposed ranking methodology to analyse the Spanish hotel industry. Spain was one of the world's favourite touristic destinations in 2019, with 83.7 million annual travellers, of which 4.3 million were international travellers. These results have boosted the Spanish economy in recent decades, and it has the second highest income from tourism (approximately 12% of its gross domestic product).

The guidelines of the United Nations World Tourism Organization (UNWTO, https://www.unwto.org/) seek a more efficient and sustainable form of tourism. Among other policies, joining resources from different touristic areas may be considered for the development of a common network that minimizes the impacts of economic fluctuations and favours the arrival of new visitors. In this sense, directly aggregating the existing resources for a set of tourism areas is naturally assumed to have the purpose of increasing the competitiveness and efficiency as a single joint destination relative to other destinations in the global tourism market.

Table 11 depicts a measure of the annual capabilities of the hotel industry in Spain per region in 2019. This dataset refers to the annual monthly average number of hotels (not including camping, rural tourism accommodations or hostels), available bed places, volume of employees, total number of hotel guests, overnight stays and estimated

Table 11
Average of monthly indicators of the Spanish hotel industry in 2019. Source: *Instituto Nacional de Estadística, INE*, https://www.ine.es/.

Region	Hotels	Bed places	Employees	Hotel guests	Overnight stays	Occupied bed places
Andalucía	2455.000	259 275.333	36 287.250	1 653 402.167	4 576 537.92	146 765.583
Aragón	766.583	37 562.333	3638.750	247 598.833	480 469.00	15 559.583
Principado de Asturias	547.833	24 187.417	2753.417	148 125.000	311 687.17	10 095.417
Illes Balears	773.667	202 184.750	33 402.917	879 083.000	4 840 793.17	152 446.583
Canarias	540.667	249 976.167	47 752.417	814 067.250	5 604 536.50	180 138.167
Cantabria	325.750	15 942.250	2157.083	110 174.583	246 587.50	7893.833
Castilla y León	1269.833	58 019.333	6710.500	425 852.250	710 568.25	23 119.083
Castilla-La Mancha	728.667	31 638.500	2818.250	189 637.500	319 173.25	10 395.667
Cataluña	2302.667	251 134.000	34 896.000	1 727 974.917	4 854 300.92	153 449.667
Comunitat Valenciana	1028.750	128 020.750	16 301.167	777 374.167	2 487 222.00	77 899.167
Extremadura	370.000	18 523.667	2350.417	122 368.250	209 967.83	6841.667
Galicia	1416.667	59 754.333	6920.917	372 374.750	751 117.92	24 260.083
Comunidad de Madrid	1170.500	114 306.500	14 758.000	1 052 319.083	2 127 934.25	69 282.500
Región de Murcia	161.000	17 745.417	2184.333	113 575.417	273 159.33	8798.833
Com. Foral de Navara	271.917	11 899.500	1492.000	89 055.083	167 078.75	5383.333
País Vasco	554.500	29 519.583	4161.500	266 280.250	513 775.25	16 609.500
La Rioja	147.500	6210.917	785.250	47 858.333	83 548.83	2718.167
Ceuta	12.667	773.500	124.667	6371.083	14 058.17	459.083
Melilla	10.000	838.000	152.000	5604.833	11 668.33	383.167

occupied bed places in 2019 for the hotel industry in the seventeen autonomous regions and two autonomous cities within the territory of Spain.

Thus, the elements of the associated DEA problem are formally defined as follows. From Table 11, the set of involved agents is given by the 17 regions and the two autonomous cities of Spain, that is, $N = \{1, \dots, 19\}$. The inputs are given by the number of hotels, occupied bed places and employees of the hotel industry. These quantities measure the hotel industry's potential in Spain. The outputs, as result of managing the existing resources, are the average of the number of hotel guests, of the overnight stays and of the occupied accommodations. Each region manages administrative hotel competencies, such as business licences, and seeks to promote new visitor arrivals through active policies in this field. Hence, a natural question refers to how efficient each of the regions is in comparison to the others. The individual efficiencies of the autonomous regions and cities under the DEA perspective are depicted in Table 12. Recall that those autonomous regions that are indistinguishable with the same relative efficiency score will occupy the artificial position given by the average position. The case of Cataluña is noteworthy because although it has the largest average amount of hotel guests, it is not efficient and occupies position 7 in the ranking of individual efficiencies.

Table 12
Individual efficiencies under a DEA approach.

Region	Efficiency	Rank	Region	Efficiency	Rank
Andalucía	0.86858	11	Extremadura	0.73014	19
Aragón	0.95428	6	Galicia	0.75457	17
Principado de Asturias	0.78051	16	Comunidad de Madrid	1.00000	2.5
Illes Balears	1.00000	2.5	Región de Murcia	0.85143	13
Canarias	1.00000	2.5	Com. Foral de Navarra	0.83709	14
Cantabria	0.80268	15	País Vasco	0.97983	5
Castilla y León	0.88999	10	La Rioja	0.85473	12
Castilla - La Mancha	0.94368	9	Ceuta	0.95249	8
Cataluña	0.95315	7	Melilla	0.74557	18
Comunitat Valenciana	1.00000	2.5			

Obviously, the leading indicators in Table 11 were drastically reduced after March 2020, although they still realistically describe the capabilities of the Spanish hotel sector. As mentioned, a long-term solution for the regions may be to work jointly to increase competitiveness and efficiency in a declining hotel market. Criteria such as geographical proximity or similar measures for promoting their destinations justify a merger of Spanish regions to act more efficiently in the global hotel market. The union of regions naturally implies the availability of all existing resources for the whole merger. The relative efficiency score of such a merger does not depend exclusively on its members because the organizational structure of the agents outside the group also influences this magnitude. Based on its ability to increase the overall efficiency score in the event of a merger, we rank the 17 autonomous regions and the 2 autonomous cities of Spain using DEA sum games.

Regarding the sampling approaches, we have used the sampling fractions in Table 13 for those coalitions $S \subseteq N$ such that $|S| \leq 11$ under the approach of Albizuri et al.. These values ensure the extraction of more than 950 partitions for each coalition and the theoretical errors that were obtained using Inequality (A.2.1) in the Online Resource Section. Otherwise, we are exhaustive in the population and consider the overall set of partitions as a sample.

Table 13
Sampling fractions, sample sizes and theoretical errors per coalition for estimating $(N, \bar{e})$.

$\|S\|$	1	2	3	4	5	6
f_S	$1.5 \cdot 10^{-9}$	$1.5 \cdot 10^{-8}$	10^{-7}	10^{-6}	$5 \cdot 10^{-6}$	$4 \cdot 10^{-5}$
ℓ_S	1024	1243	1049	1383	955	1106
ε	0.04244	0.03852	0.04193	0.03652	0.04395	0.04084
$\|S\|$	7	8	9	10	11	
f_S	$2.5 \cdot 10^{-4}$	$1.5 \cdot 10^{-3}$	10^{-2}	$5 \cdot 10^{-2}$	0.250	
ℓ_S	1054	1018	1160	1058	1035	
ε	0.04183	0.04257	0.03988	0.04175	0.04222	

Under the approach of Hu and Yang, we use $f = 10^{-10}$ for each $S \subseteq N$, which results in a sample size of 584 partitions per coalition. Inequality (A.1.1) in the Online Resource Section ensures a theoretical error in estimating $\bar{\bar{e}}$ equal to $\varepsilon = 0.0562$ for this sample size when $\alpha = 0.05$.

Columns 3 and 5 in Table 14 depicts the rankings based on the Shapley value of the TU games of de Clippel and Serrano (2008) (CS) and McQuillin (2009) (MQ), and columns 7 and 9 show the rankings based on the estimated TU games of Albizuri et al. (2005) (A) and of Hu and Yang (2010) (HY), respectively.

Based on the results, we draw some conclusions about the influence of the regions involved on the overall efficiency score in the case of a merger. Illes Balears occupies the first position under the four considered approaches. Madrid is always in second position in the rankings, and the Shapley value for the TU games of (MQ), (A), and (HY) assign the fourth largest component to Comunitat Valenciana. Under (CS), this region moves up to third. Canarias falls to sixth position under the approaches of (CS), of (A), and of (HY), whereas it is last in the ranking obtained under (MQ). Cataluña usually ranks third except for (CS), where it moves down to fourth place. País Vasco is in the fifth position except under the approach of (MQ), which assigns it sixth place. Ceuta occupies positions 8 and 5 under (CS) and (MQ), respectively, but position 7 under (A) and (HY). Melilla occupies the fifteenth position under (CS) and (HY), the seventh under (MQ) and the fourteenth under (A). Andalucía ranks eleventh under (MQ), fourteenth under (CS) and (HY), and fifteenth under (A). The same

Table 14
Rankings under the approaches of de Clippel and Serrano (CS) and McQuillin (MQ), and of Albizuri et al. (A) and Hu and Yang (HY) with sampling.

Region	Exact rankings				Estimated rankings			
	(CS)	Rank	(MQ)	Rank	(A)	Rank	(HY)	Rank
Andalucía	0.04017	14	0.05232	11	0.04237	15	0.04133	14
Aragón	0.05117	7	0.05181	15	0.05287	8	0.05202	8
Principado de Asturias	0.03417	17	0.05194	14	0.03704	17	0.03606	17
Illes Balears	0.10943	1	0.05596	1	0.09270	1	0.09902	1
Canarias	0.05665	6	0.05031	19	0.05545	6	0.05620	6
Cantabria	0.03790	16	0.05223	12	0.04122	16	0.03985	16
Castilla y León	0.04094	13	0.05131	17	0.04371	13	0.04266	13
Castilla - La Mancha	0.04737	9	0.05156	16	0.04882	9	0.04791	9
Cataluña	0.07229	4	0.05518	3	0.07094	3	0.07228	3
Comunitat Valenciana	0.07726	3	0.05389	4	0.07076	4	0.07227	4
Extremadura	0.03001	18	0.05205	13	0.03341	18	0.03225	18
Galicia	0.02533	19	0.05094	18	0.02817	19	0.02707	19
Comunidad de Madrid	0.10352	2	0.05537	2	0.08920	2	0.09413	2
Región de Murcia	0.04227	11	0.05243	9	0.04697	11	0.04587	11
Com. Foral de Navarra	0.04212	12	0.05234	10	0.04543	12	0.04426	12
País Vasco	0.05668	5	0.05262	6	0.05802	5	0.05753	5
La Rioja	0.04391	10	0.05245	8	0.04718	10	0.04607	10
Ceuta	0.05090	8	0.05263	5	0.05306	7	0.05214	7
Melilla	0.03793	15	0.05261	7	0.04271	14	0.04109	15

scheme is repeated in Murcia, Navarra and La Rioja. Murcia ranks ninth under (MQ) and eleventh under (CS), (A) or (HY). Navarra occupies position 10 under (MQ) and falls to position 12 under the remaining approaches. Moreover, La Rioja occupies position 8 under (MQ) and position 10 under (CS), (A) and (HY). Aragón occupies position 7 under (CS), position 8 under (A) and (HY), and position 15 under (MQ). Asturias occupies position 14 under (MQ) and position 17 under (CS), (A), and (HY). Cantabria, Castilla *y* León and Castilla - La Mancha rank sixteenth, thirteen and ninth for all approaches, respectively, except for the (MQ) approach, under which they move to positions 12, 17, and 16, respectively. Finally, we consider the cases of Extremadura and Galicia, which occupy positions 18 and 19 under (CS), (A), and (HY), respectively, and rise to positions 13 and 18 under (MQ), respectively. Broadly speaking, the ranking based on (CS) and those obtained from the Shapley value estimations for the TU games of (A) and (HY) are quite similar. The biggest differences among the rankings occur under the (MQ) approach. These conclusions coincide with those derived from the Spearman's rank correlation matrix in Table 15.

Table 15
Spearman's correlation matrix for the rankings of the hotel industry in Spain.

	(CS)	(MQ)	(A)	(HY)
(CS)	1.00000	0.60877	0.99474	0.99649
(MQ)	0.60877	1.00000	0.63509	0.62807
(A)	0.99474	0.63509	1.00000	0.99825
(HY)	0.99649	0.62807	0.99825	1.00000

7. Concluding remarks

The data envelopment analysis (DEA) method has received considerable research attention in recent decades due to its multiple applications. A wide variety of papers has focused on ranking DMUs according to their relative efficiency scores by distinguishing efficient DMUs from nonefficient DMUs. In this work, we have analysed the impact of the direct aggregation of the inputs and the outputs of DMUs on the overall efficiency score in the event of a merger. In this framework, this worth can be influenced by an organization of agents outside of the coalition because all DMUs are considered regardless of the value of their individual efficiencies. Thus, partition function form games (Thrall & Lucas, 1963) are considered to model these situations.

The theoretical properties satisfied by the new classes of games are formally studied. We use some values defined for games in partition function form to rank DMUs. Specifically, we consider the proposals of Albizuri et al. (2005), de Clippel and Serrano (2008), Hu and Yang (2010), and McQuillin (2009), which are based on the Shapley value of specific TU games. Due to the difficulties in exactly determining some of these values, we propose two sampling proposals as an alternative in these challenging situations. As an application, we use these proposed methods in the analysis of the hotel sector in Spain in 2019. In accordance with the country's territorial organization, cooperation among Spanish regions aims to improve of the overall efficiency score.

As mentioned, our proposal is formulated in terms of the output-oriented DEA model with constant returns to scale. Further research should address a comparative study that includes the performance of alternative DMU ranking methodologies, such as the ones considered in Labijak-Kowalska and Kadzinski (2021), and several DEA models, such as those with variable return to scale. In addition, the consideration of alternative values in the context of partition function form games may be of interest for the task of ranking DMUs. From a computational perspective, the use of alternative sampling methodologies for this purpose can be addressed.

CRediT authorship contribution statement

A. Saavedra-Nieves: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. **M.G. Fiestras-Janeiro:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work has been supported by FEDER/*Ministerio de Ciencia, Innovación y Universidades - Agencia Estatal de Investigación*, Spain under grants MTM2017-87197-C3-2-P and MTM2017-87197-C3-3-P, and by the Xunta de Galicia through the European Regional Development Fund (*Grupos de Referencia Competitiva* ED431C-2017/38) and by the *Consellería de Cultura, Educación e Universidades, Xunta de Galicia*, Spain (*Grupos de Referencia Competitiva* ED431C-2020/03).

Authors also acknowledge Professor Ignacio García-Jurado for his insightful comments on an earlier version of this paper.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.eswa.2022.116571. To facilitate the readability of the paper, the statistical results in Section 5 are reported in the Online Resource Section provided with this paper.

References

Adler, N., Friedman, L., & Sinuany-Stern, Z. (2002). Review of ranking methods in the data envelopment analysis context. *European Journal of Operational Research*, *140*, 249–265.

Albizuri, M. J., Arin, J., & Rubio, J. (2005). An axiom system for a value for games in partition function form. *International Game Theory Review*, *7*, 63–72.

An, Q., Wen, Y., Ding, T., & Li, Y. (2019). Resource sharing and payoff allocation in a three-stage system: Integrating network DEA with the Shapley value method. *Omega*, *85*, 16–25.

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

Banker, R. D., Cooper, W. W., Seiford, L. M., & Zhu, J. (2011). Returns to scale in DEA. In *Handbook on data envelopment analysis* (pp. 41–70). Springer.

Castro, J., Gómez, D., & Tejada, J. (2009). Polynomial calculation of the Shapley value based on sampling. *Computers & Operations Research*, *36*, 1726–1730.

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

Chu, J., Wu, J., Chu, C., & Zhang, T. (2020). DEA-based fixed cost allocation in two-stage systems: Leader-follower and satisfaction degree bargaining game approaches. *Omega*, *94*, Article 102054.

Cochran, W. G. (2007). *Sampling techniques*. John Wiley & Sons.

Cook, W. D., & Kress, M. (1999). Characterizing an equitable allocation of shared costs: A DEA approach. *European Journal of Operational Research*, *119*, 652–661.

Cooper, W. W., Ruiz, J. L., & Sirvent, I. (2009). Selecting non-zero weights to evaluate effectiveness of basketball players with DEA. *European Journal of Operational Research*, *195*, 563–574.

Csercsik, D., & Kóczy, L. A. (2017). Efficiency and stability in electrical power transmission networks: A partition function form approach. *Networks and Spatial Economics*, *17*, 1161–1184.

de Clippel, G., & Serrano, R. (2008). Marginal contributions and externalities in the value. *Econometrica*, *76*, 1413–1436.

Demetrescu, C., Lupia, F., Mendicelli, A., Ribichini, A., Scarcello, F., & Schaerf, M. (2019). On the Shapley value and its application to the Italian VQR research assessment exercise. *Journal of Informetrics*, *13*, 87–104.

Fernández-García, F. R., & Puerto-Albandoz, J. (2006). *Teoría de juegos multiobjetivo*. Sevilla: Imagraf Impresores SA.

González-Díaz, J., García-Jurado, I., & Fiestras-Janeiro, M. G. (2010). *An introductory course on mathematical game theory, Vol. 115*. Providence: American Mathematical Society.

Hafalir, I. E. (2007). Efficiency in coalition games with externalities. *Games and Economic Behavior*, *61*, 242–258.

Hinojosa, M. A., Lozano, S., Borrero, D. V., & Mármol, A. M. (2017). Ranking efficient DMUs using cooperative game theory. *Expert Systems with Applications*, *80*, 273–283.

Hosseinzadeh Lotfi, F., Noora, A. A., Jahanshahloo, G. R., & Reshadi, M. (2011). One DEA ranking method based on applying aggregate units. *Expert Systems with Applications*, *38*, 13468–13471.

Hu, C.-C., & Yang, Y.-Y. (2010). An axiomatic characterization of a value for games in partition function form. *SERIEs*, *1*, 475–487.

Hua, Z., Bian, Y., & Liang, L. (2007). Eco-efficiency analysis of paper mills along the Huai River: An extended DEA approach. *Omega*, *35*, 578–587.

Kritikos, M. N. (2017). A full ranking methodology in data envelopment analysis based on a set of dummy decision making units. *Expert Systems with Applications*, *77*, 211–225.

Labijak-Kowalska, A., & Kadzinski, M. (2021). Experimental comparison of results provided by ranking methods in Data Envelopment Analysis. *Expert Systems with Applications*, *173*, Article 114739.

Li, Y., & Liang, L. (2010). A Shapley value index on the importance of variables in DEA models. *Expert Systems with Applications*, *37*, 6287–6292.

Li, Y., Xie, J., Wang, M., & Liang, L. (2016). Super efficiency evaluation using a common platform on a cooperative game. *European Journal of Operational Research*, *255*, 884–892.

Liu, X., Lindroos, M., & Sandal, L. (2016). Sharing a fish stock when distribution and harvest costs are density dependent. *Environmental and Resource Economics*, *63*, 665–686.

Lozano, S. (2012). Information sharing in DEA: A cooperative game theory approach. *European Journal of Operational Research*, *222*, 558–565.

Lozano, S., Hinojosa, M. A., Mármol, A. M., & Borrero, D. V. (2016). DEA and cooperative game theory. In *Handbook of operations analytics using data envelopment analysis* (pp. 215–239). Springer.

Mahdiraji, H. A., Razghandi, E., & Hatami-Marbini, A. (2021). Overlapping coalition formation in game theory: A state-of-the-art review. *Expert Systems with Applications*, *174*, Article 114752.

Maleki, S. (2015). *Addressing the computational issues of the Shapley value with applications in the smart grid* (Unpublished doctoral dissertation), University of Southampton.

McQuillin, B. (2009). The extended and generalized Shapley value: Simultaneous consideration of coalitional externalities and coalitional structure. *Journal of Economic Theory*, *144*, 696–721.

Omrani, H., Shafaat, K., & Emrouznejad, A. (2018). An integrated fuzzy clustering cooperative game data envelopment analysis model with application in hospital efficiency. *Expert Systems with Applications*, *114*, 615–628.

Pham Do, K. H., & Norde, H. (2007). The Shapley value for partition function form games. *International Game Theory Review*, *9*, 353–360.

Pintassilgo, P., & Lindroos, M. (2008). Application of partition function games to the management of straddling fish stocks. In *Game theory and policymaking in natural resources and the environment* (pp. 65–85). London, New York: Routledge.

R Core Team (2021). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, Retrieved from https://www.R-project.org/.

Saavedra-Nieves, A., García-Jurado, I., & Fiestras-Janeiro, M. G. (2018). Estimation of the owen value based on sampling. In E. Gil, E. Gil, J. Gil, & M. A. Gil (Eds.), *The mathematics of the uncertain: A tribute to Pedro Gil* (pp. 347–356). Springer.

Shapley, L. S. (1953). A value for n-person games. *Contributions to the Theory of Games*, *2*(28), 307–317.

Thrall, R. M., & Lucas, W. F. (1963). N-person games in partition function form. *Naval Research Logistics Quarterly*, *10*, 281–298.

Yang, G., Sun, H., Hou, D., & Xu, G. (2019). Games in sequencing situations with externalities. *European Journal of Operational Research*, *278*, 699–708.

Yang, Z., & Zhang, Q. (2015). Resource allocation based on DEA and modified Shapley value. *Applied Mathematics and Computation*, *263*, 280–286.

Yazdani, M., Chatterjee, P., Pamucar, D., & Chakraborty, S. (2020). Development of an integrated decision making model for location selection of logistics centers in the spanish autonomous communities. *Expert Systems with Applications*, *148*, Article 113208.

Yin, P., Chu, J., Wu, J., Ding, J., Yang, M., & Wang, Y. (2020). A DEA-based two-stage network approach for hotel performance analysis: An internal cooperation perspective. *Omega*, *93*, Article 102035.