Incentivizing Centrally Regulated Units to Improve Performance: Pitfalls and Requirements

Mohsen Afsharian[1,3], Heinz Ahn[2], Emmanuel Thanassoulis[4,5]

Corresponding author (Mohsen Afsharian; m.afsharian@tu-bs.de)

8/

[1] *Chair of Production and Logistics, Leibniz-FH, University of Applied Sciences, Expo Plaza 11, 30539, Hanover, Germany*

[2] *Department of Business Sciences, Technische Universität Braunschweig, Fallersleber-Tor-Wall 23, 38100 Braunschweig, Germany*

[3] *Florida International University, Department of Information Systems and Business Analytics 11200 SW 8th Street, Miami, Florida 33199, U.S.A.*

[4] *Department of Operations and Information Management, Aston Business School, Aston University, Birmingham B4 7ET, UK*

[5] *Associate Oxera LLP, Oxford, UK*

Abstract. This paper explores the foundations for developing incentives for influencing units operating within centrally managed organizations. We begin by laying out the theory of managerial control in principal-agent contexts and draw from the incentive mechanisms developed in the related field of economic regulation. In particular, we highlight issues, differences and similarities in three recently proposed approaches under these circumstances, not only to compare them, but more importantly to motivate and arrive at requirements that should be met by incentivization systems in centrally managed multi-unit organizations. The stipulated requirements are not intended to be exhaustive but rather aim at defining conceptual foundations for further discussions and encouraging avenues for future research in this field. Our investigations are supported by graphical examples and an analysis of empirical data from banking.

Keywords: Data Envelopment Analysis; Central Management; Incentive Regulation; Incentivization Mechanism

1. Introduction

In applications of Data Envelopment Analysis (DEA), there are situations where a central body manages a set of decision making units (DMUs) delivering some services. Examples are a supermarket chain managing its stores, a fast-food restaurant chain managing its outlets, a bank managing its branches etc.[1] In such multi-unit organizations, the central management desires a mechanism by which the local management of each unit is incentivized to perform towards the improvement of the performance of the organization as a whole (rather than solely to its own benefit). In such instances, there is a common understanding that units found to be inefficient should be encouraged to make efficiency savings. On the other hand, units that are identified to be efficient should be incentivized by a reward consistent with the level of their impact on the efficiency of the system of units as a whole (for more details see Varmaz et al. 2013; Afsharian et al. 2017; Fang 2020).

Varmaz et al. (2013) were the first to propose a DEA-based incentives system under these circumstances – we refer to it as Varmaz et al.'s approach. In this approach, it is assumed that the central decision maker aims to minimize the overall input consumption by the units given the aggregated outputs they produce. To operationalize the system, the centralized DEA program proposed by Lozano and Villa (2004) has been used in a modified incentivization mechanism. The original framework of this mechanism (we refer to it as "decentralized system of incentives") applies conventional DEA models, i.e., those where the aim is to improve the performance of each unit independently (for more details, see Bogetoft 2013). Varmaz et al. (2013) adapted this framework in such a way that it yields a "super-efficiency" measure for each unit in turn, to be used to set individualized incentives under central management.

Afsharian et al. (2017) identified shortcomings in Varmaz et al.'s approach and used it as a starting point to develop a new DEA-based system for incentivizing operating units under central management – we refer to their system as Afsharian et al.'s approach. More precisely, the super-efficiency measure in Varmaz et al.'s approach is redefined and mathematical programs are further developed to capture more accurately how the efficient frontier of the system of units varies with and without each unit under consideration. This leads to a more appropriate level of incentives for each unit under central management. Furthermore, it is shown that in such systems, units can be "joint super-efficient" in terms of input-output levels. An adaptation of the sensitivity-based procedure introduced by Thanassoulis (1999) is proposed so that units that are jointly super-efficient can have their individual super-efficiency identified.

Fang (2020, p. 158) has recently claimed that "Afsharian et al.'s (2017) approach can overcome the shortcomings underling [sic!] Varmaz et al.'s (2013) model, while it cannot measure the individual's contribution of each unit to the efficiency of the whole system appropriately and comprehensively". The

[1] In a broader context, an appropriate modification of the approaches being discussed in this paper may be used in cases in which there exist natural monopolies instead of usual competitive markets. Examples are those of large infrastructure industries like water, electricity and gas networks. For a comprehensive overview see, e.g., Bogetoft (2013); Agrell and Bogetoft (2017); Agrell and Bogetoft (2018); Afsharian et al. (2022).

author has therefore proposed an alternative to Afsharian et al.'s approach to overcome the identified issues. We refer to this approach as Fang's approach. In the core of this system of incentives, Fang (2020) suggests the frontier change (similar to the one in Afsharian et al.'s approach) and change in the technical efficiency as two sources of impact on the efficiency of the whole system of units. In particular, the latter captures how the efficiency score of the so-called aggregate unit – representing the whole system of units (yet to be defined precisely) – varies with and without the inputs and outputs of the unit under consideration.

We explore the above measures of super-efficiency as a means for identifying units of centralized multi-unit organizations that should be incentivized to reveal efficient operating practices. This would raise benchmark performance for constituent units and so for the organization as a whole. In this context, we set out requirements that should be met by incentivization approaches based on (super-)efficiency measures.

Against this background, the contributions of this paper are organized as follows:

- In Section 2, we formalize the problem setting of how the structure of the above-outlined centrally managed multi-group organizations is modelled within the context of the principal-agent framework in general and that of DEA-based incentive mechanisms in particular.
- In Section 3, a technical overview of the use of DEA for incentivizing units under both decentralized and central management is given. In the latter case – which is the focus of this paper – the key features of the aforementioned three systems of incentives are exposed.[2]
- In Section 4, we highlight issues, differences and similarities in these three approaches, not only to compare them, but more importantly to motivate and arrive at essential requirements that should be met by incentivization approaches in centrally managed multi-unit organizations.
- As we conclude in Section 5, the stipulated requirements are not intended to be exhaustive but rather aim at defining conceptual foundations for further discussions and encouraging avenues for future research in this field.
- Our investigations are supported by graphical examples and an analysis of empirical data from banking.
- We use the corresponding data set for a step-by-step illustration and comparison between the above-mentioned three approaches. As these approaches have applied the same data set, we are enabled to refer to the statements and interpretations given by the respective authors, if necessary.

[2] Note that these approaches have also been further developed for other purposes of incentivization in centrally managed multi-unit organizations. For example, Afsharian et al.'s approach is used as a basis to design a system of incentives for units that are organized into a few distinct management groups (see Afsharian 2020). More recently, it was also suggested that radial DEA models in Fang's approach should be replaced with slack-based models to more appropriately capture the efficiency scores (Davtalab-Olyaie et al. 2021). There are also other methods for the sake of incentivization (see, e.g., Afsharian et al. 2019 and Dai 2021). A comprehensive consideration of such methods is not pursued here as they are not built upon the three approaches discussed in this paper.

2. Problem setting and fundamentals

A management control system is typically defined as a systematic organizational process by which the resources are ensured to be strategically obtained and used efficiently in line with the organizational strategies and in ways that lead to the attainment of organizational objectives (Anthony et al. 2014). On this basis, regarding the typology suggested by Malmi and Brown (2008), a management control system can contain a package of control processes and mechanisms which are run in the organization in order to accomplish a set of pre-specified goals. Examples of such mechanisms are planning controls (Flamholtz et al. 1985), administrative controls (Simons 1987), cultural controls (Birnberg and Snodgrass 1988), cybernetic controls (Green and Welsh 1988) as well as reward and compensation controls (Bonner and Sprinkle 2002). For a detailed discussion, see also Anthony et al. (2014) and Afsharian (2022).

A narrower definition of the management control system being applied in intra-organizational or centralized management refers to the systematic process by which the organization's higher-level managers (i.e., central decision maker) influence the organization's lower-level managers (i.e., local decision makers) in order to implement the strategies and to pursue mutual goals (Flamholtz et al. 1985; Pernot and Roodhooft 2014). However, such a centralized framework does not often allocate all the power to make decisions that affect the future of the organization to the central decision maker. Some of this power is shared with the local decision makers (i.e., DMUs under central management) who are responsible for controlling their local variables. Due of this flexibility in such a framework, it is often realized that local managers in the organization do not automatically perform actions that are imposed from above (Afsharian and Ahn 2017). Hence, management control systems typically include an appropriate incentivization mechanism with the aim to ensure that all processes and activities, on which local managers are in charge of, will create the desired future of the organization. We refer to Afsharian (2022) for a detailed discussion on this topic.

In the context of the above, there is a tight relation between the results of the efficiency measurement systems and the incentivization mechanism in management control systems for improving performance. This correspondence can be implicit or explicit (see, e.g., Agrell et al. 2002; Bogetoft 2013). In the former case, it has been shown that having even solely an efficiency measurement system associated with an appropriate reporting mechanism can improve the performance, as the DMUs under evaluation pay thoughtful attention to it. An explicit relationship can be defined by means of innovative incentive methods, which specify the budgeting rules, the salary plans, the tariff regulations, etc. In particular, an incentive method can be defined by which the responsible employees in the operating units are rewarded on the basis of the results from the efficiency measurement systems. It has been shown that the existence of such a system can improve the performance significantly (for a detailed discussion on this topic, see also Afsharian 2022 and Thanassoulis et al. 2022 for the case of regulated industries).

In the case of an explicit approach, the problem of incentivizing operating units to improve performance can be modelled in a principal-agent context. The principal (e.g., central management) does not have access to full information as to the true cost function that pertains to each agent (unit) in delivering the outputs demanded by the principal. This leads to an asymmetry of information, which can be exploited by the agents to extract rents, i.e., it takes effort for the agents to be cost-efficient and so they tend to slacken effort (extract rent). The principal wishes to reduce this rent by incentivizing the agents to reveal information that leads to cost-efficient behaviour by them (for more details see, e.g., Bogetoft 2013; Agrell and Bogetoft 2017).

Let us assume that there exists a set of *n* DMUs, which use a set of *m* inputs to deliver a set of *s* outputs. Let $X_j = (x_{1j}, x_{2j}, \ldots, x_{mj}) \in \Re_+^m$ and $Y_j = (y_{1j}, y_{2j}, \ldots, y_{sj}) \in \Re_+^s$ be nonnegative and nonzero vectors of inputs and outputs of DMU_j ($j = 1, \ldots, n$).

Let DMU_p be the unit under evaluation. Based upon the work of Shleifer (1985) – Bogetoft (1997) introduced the following frontier-based yardstick formula through which a so-called compensation plan (e.g., a budget or a monetary transfer from the central management to the agent) for DMU_p (indicated by b_p) in a system of incentives can be determined:

$$b_p = c_p + \rho(\theta_p - 1)\, c_p. \tag{1}$$

In this formula, c_p is the cost of the operations of DMU_p and θ_p quantifies its efficiency, which can be computed by an appropriate DEA program using the observed input-output data. On this basis, (θ_p –1) is the fraction of c_p available for saving. We note here that $0 < \rho < 1$ is a subjective parameter which can moderate the savings fraction (θ_p –1) imposed on DMU_p. (for more details, see, e.g., Agrell et al. 2005). On this basis, the costs of efficient units are compensated completely whereas the inefficient units are not fully compensated but are obliged to save costs in line with the formula in (1).

It is essential to see why the above mechanism encourage DMU_p to operate as efficiently as possible. Following Bogetoft (1997), we assume that each DMU_p seeks to maximize its utility defined as the weighted sum of its profit and slack (i.e., excess costs) as follows (see also Bogetoft 2000):

$$U_p = \underbrace{\left(b_p - c_p\right)}_{profit} + \rho \underbrace{\left(c_p - c_p^{DEA}\right)}_{slack}, \tag{2}$$

where, c_p and b_p are those defined above. In this formula, $c_p^{DEA} = \theta_p c_p$ represent the so-called efficient costs of operations of DMU_p. As will be seen in the next section, different ways have been suggested (within the three systems of incentives being reviewed in this paper) to compute θ_p and accordingly c_p^{DEA}. On this basis, $b_p - c_p$ and $c_p - c_p^{DEA}$ represent the profit and slack for DMU_p, respectively. Furthermore, the parameter ρ specifies the value of the slack relative to the profit. Hence, as far as the slack is weighted less than the profit, the mechanism induces each DMU$_p$ to minimize its slack for maximizing its utility. Bogetoft (1997, p. 283) argues that the agents "must get a utility of at least 0. It ensures that all DMUs will accept the reimbursement plans taking into account their superior information". Hence, if we let the DMUs' reservation utilities be 0 and replace c_p^{DEA} with $\theta_p c_p$ in (2), the resulting compensation plan for DMU$_p$ will be the one in (1). More discussions on this topic can be found in Bogetoft (2013).

We should also note that the system of incentives of Bogetoft (1997) has been designed for networks of public service providers and franchised monopolies that do not run in a competitive market. Nevertheless, according to Bogetoft (1997, p. 283) "alternatively, one may think of (public or private) organizations within the educational or health sectors, where consumers' costs of service are covered to a large extent by the State or by privately held insurances. One may also think of large private organizations like chain stores, banks with several branch offices or fast food companies with many outlets, that produce the same spectrum of homogeneous goods which are primarily marketed by the company in large". Hence, an appropriate modification of the system of incentives proposed by Bogetoft (1997) can be used under central management if the characteristics of the network of the units lead to a market structure – as outlined above – in which the tendency towards monopoly is promoted.

3. Technical overview of the approaches

Under the so-called decentralized systems of incentives (see, e.g., Bogetoft 1997), the efficiency of a DMU$_p$ under evaluation has mainly been measured by conventional DEA programs, such as the following one (Banker et al. 1984):

$$\theta_p = \max_{v_{ip}, u_{rp}, \mu_p} \left\{ \sum_{r=1}^{s} u_{rp} y_{rp} + \mu_p \;\middle|\; \begin{array}{ll} \sum_{i=1}^{m} v_{ip} x_{ip} = 1, & \\ \sum_{r=1}^{s} u_{rp} y_{rj} - \sum_{i=1}^{m} v_{ip} x_{ij} + \mu_p \le 0, & j = 1,\dots,n \\ v_{ip} \ge 0,\; i = 1,\dots,m,\; u_{rp} \ge 0,\; r = 1,\dots,s,\; \mu_p \; \textit{free in sign} & \end{array} \right\}. \tag{3}$$

In this program, v_{ip} and u_{rp} are the weights of inputs and outputs, respectively, and they are the variables in the model. θ_p represents the relative efficiency score of DMU$_p$ and μ_p reflects the scale size of this unit (for a full discussion of conventional DEA programs and their features, we refer to Thanassoulis 2001).

As a fundamental objective in systems of incentives, the central management seeks to incentivize the agents to perform better by minimizing their operating costs. One means at the disposal of the central management is the revealed information about the operating costs of the agents. By using frontier methods such as DEA, it is possible to reveal the target efficient costs for each agent (see also the discussion in Section 2). The agent is then induced to move towards (not necessarily to) the efficient cost level.

However, this approach may not be suitable as an incentive mechanism for units that are already efficient. They will have no incentive to reveal further efficient cost norms to be incorporated into the compensation formula, leading to the so-called ratchet effect (see, e.g., Agrell et al. 2005; Bottasso and Conti 2009). For such cases, i.e., in order to incentivize already efficient units, Bogetoft (1997) suggested that the super-efficiency variation of the program in (3) – introduced by Andersen and Petersen (1993) – be applied. This can be done by running the program in (3) under the assumption that DMU$_p$ under evaluation is excluded from the reference set (from the set of inequality constraints):

$$\theta_p^s = \max_{v_{ip}, u_{rp}, \mu_p} \left\{ \sum_{r=1}^{s} u_{rp} y_{rp} + \mu_p \;\middle|\; \begin{array}{l} \sum_{i=1}^{m} v_{ip} x_{ip} = 1, \\ \sum_{r=1}^{s} u_{rp} y_{rj} - \sum_{i=1}^{m} v_{ip} x_{ij} + \mu_p \le 0, \qquad j = 1,\ldots,n, \quad j \ne p \\ v_{ip} \ge 0, \; i = 1,\ldots,m, \; u_{rp} \ge 0, \quad r = 1,\ldots,s, \; \mu_p \textit{ free in sign} \end{array} \right\}. \tag{4}$$

With an application of (4), we are then able to capture the movement of the frontier when DMU$_p$ is dropped from the set of potential benchmark units, which shape the efficient frontier. In the context of the incentive formula in (1), θ_p^s forms a basis for providing this unit with a reward to further push out the frontier, which could in turn lead to cost-efficient behaviour by the units in future.

There is another way to present the above procedure, which could be helpful to understand the logics behind the following three approaches under central management. According to (3) and (4), let the efficiency and super-efficiency score of DMU$_p$ under evaluation be shown by θ_p and θ_p^s, respectively. Define the following ratio:

$$\tilde{\theta}_p = \left(\theta_p^s / \theta_p\right) \times \theta_p \,. \tag{5}$$

Obviously, $\tilde{\theta}_p$ in (5) will provide the same score $\tilde{\theta}_p = \theta_p$ as in (3) if DMU$_p$ is inefficient, because in this case $\theta_p^s = \theta_p$. However, if DMU$_p$ is efficient and a benchmark (i.e., it shapes the efficient frontier), it will have a score $\tilde{\theta}_p$ greater than one, because $\theta_p^s > \theta_p$, while $\theta_p = 1$. Hence, utilizing $\tilde{\theta}_p$ in (4), units with poor performance are not fully reimbursed, while performers with super-efficiency $\tilde{\theta}_p > 1$ can be awarded budgets above c_p to incentivize them to better perform in the future.

Under central management, what is ultimately sought in the approaches of Varmaz et al. (2013), Afsharian et al. (2017) and Fang (2020), is also a type of super-efficiency of some DMU$_p$ to be applied within an incentive system to induce the units to reveal more efficient practices. In all three approaches, it is established that units found to be inefficient should be encouraged to make efficiency savings. On the other hand, units that are identified to be efficient need to be incentivized by a reward consistent with the level of their impact on the system's overall efficiency, overcoming the ratchet effect in the system of incentives.

In order to measure the super-efficiency of each unit under central management, Varmaz et al.'s approach compares the efficiency of the entire system with the efficiency score of the system where the unit under evaluation is excluded from the data set. More specifically, the authors use two DEA models. First, the DEA program proposed by Lozano and Villa (2004) is used to compute a measure of aggregate efficiency under central management:

$$Eff = \max_{v_i, u_r, \mu} \left\{ \sum_{j=1}^{n}\sum_{r=1}^{s} u_r y_{rj} + n\mu \;\middle|\; \begin{array}{ll} \sum_{j=1}^{n}\sum_{i=1}^{m} v_i x_{ij} = 1, & \\ \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} + \mu \le 0, & j = 1,\dots,n \\ v_i \ge 0,\; i=1,\dots,m;\; u_r \ge 0,\; r=1,\dots,s;\; \mu \textit{ free in sign} & \end{array} \right\}. \tag{6}$$

This program determines a single overall efficiency score Eff for the whole system of units or equivalently for a grand unit that possesses the aggregate (or the average) value of inputs and outputs computed across all units in the system[3] (see the objective function and the normalization constraint; for more details on this program see, e.g., Asmild et al. 2009; Mar-Molinero et al. 2014).

In Varmaz et al.'s approach, the super-efficiency score of a DMU$_p$ – to specify further the respective compensation level for this unit within the formula in (1) – is then defined as

$$\tilde{\theta}_p^{VA} = Eff \,/\, Eff_{\neq p}^{\neq p}\,, \tag{7}$$

where $Eff_{\neq p}^{\neq p}$ is also computed by (6) with the assumption that this unit is excluded from the data set:

$$Eff_{\neq p}^{\neq p} = \max_{v_i, u_r, \mu} \left\{ \sum_{\substack{j=1\\ j\neq p}}^{n}\sum_{r=1}^{s} u_r y_{rj} + (n-1)\mu \;\middle|\; \begin{array}{ll} \sum_{\substack{j=1\\ j\neq p}}^{n}\sum_{i=1}^{m} v_i x_{ij} = 1, & \\ \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} + \mu \le 0, & j = 1,\dots,n, \quad j \neq p \\ v_i \ge 0,\; i=1,\dots,m;\; u_r \ge 0,\; r=1,\dots,s;\; \mu \textit{ free in sign} & \end{array} \right\}. \tag{8}$$

[3] As argued in Ylvinger (2000), the grand unit (or the average unit in this case) can be used to evaluate the efficiency for a system of units when a reallocation of inputs across the units is allowed. Otherwise, the use of the average unit may bias the measure of the overall efficiency. Hence, the application of the program in (6) is only advised where this requirement is met (see also the discussions in Afsharian 2021).

Afsharian et al. (2017) argue that a common set of input-output weights v_i^* and u_r^* should be derived from the program in (6). This set should first be applied to the input-output level of each unit in the system to calculate an intermediate efficiency score[4]

$$\theta_p^{int} = \frac{\sum_{r=1}^{s} u_r^* y_{rp} + \mu^*}{\sum_{i=1}^{m} v_i^* x_{ip}} . \tag{9}$$

The super-efficiency score of a DMU$_p$ is then suggested as

$$\tilde{\theta}_p^{Af} = \left(Eff_{\neq p} / Eff \right) \times \theta_p^{int} , \tag{10}$$

where $Eff_{\neq p}$ is also computed by (6) with the assumption that DMU$_p$ is excluded only from candidate boundary units:

$$\left. \begin{array}{l} Eff_{\neq p} = \max\limits_{v_i, u_r, \mu} \left\{ \sum_{j=1}^{n} \sum_{r=1}^{s} u_r y_{rj} + n\mu \;\middle|\; \sum_{j=1}^{n} \sum_{i=1}^{m} v_i x_{ij} = 1, \right. \\ \qquad \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} + \mu \le 0, \qquad j = 1,\dots,n, \quad j \neq p \\ \qquad v_i \ge 0, \;\; i = 1,\dots,m; \;\; u_r \ge 0, \;\; r = 1,\dots,s; \;\; \mu \textit{ free in sign} \end{array} \right\} . \tag{11}$$

Fang (2020) has recently proposed that the unit under evaluation should be excluded once from the candidate boundary units (where the grand unit remains unchanged) and once from the grand unit (where the boundary remains intact). More specifically, the super-efficiency score of a DMU$_p$ is defined as

$$\tilde{\theta}_p^{Fa} = Eff_{\neq p} / Eff^{\neq p} , \tag{12}$$

where $Eff_{\neq p}$ is the same as in (11), but $Eff^{\neq p}$ is computed by (6) with the assumption that DMU$_p$ is excluded only from the grand unit (i.e., from the objective function and the normalization constraint):

$$\left. \begin{array}{l} Eff^{\neq p} = \max\limits_{v_i, u_r, \mu} \left\{ \sum_{\substack{j=1 \\ j \neq p}}^{n} \sum_{r=1}^{s} u_r y_{rj} + (n-1)\mu \;\middle|\; \sum_{\substack{j=1 \\ j \neq p}}^{n} \sum_{i=1}^{m} v_i x_{ij} = 1, \right. \\ \qquad \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} + \mu \le 0, \qquad j = 1,\dots,n \\ \qquad v_i \ge 0, \;\; i = 1,\dots,m; \;\; u_r \ge 0, \;\; r = 1,\dots,s; \;\; \mu \textit{ free in sign} \end{array} \right\} . \tag{13}$$

[4] An extended program was also proposed in Afsharian et al. (2017) to result in a unique optimal solution with unique efficiency scores, if multiple solutions (though rarely) exist for the grand unit.

We note that, depending on the context, one may apply other types of centralized DEA programs introduced in the literature (for a review of centralized DEA programs, see, e.g., Afsharian et al. 2021). For example, one interesting proposal has recently been given by Davtalab-Olyaie et al. (2021). Within Fang's approach, the authors suggest applying corresponding slack-based centralized DEA programs instead of the above radial programs to result in Pareto efficient solutions. Note that the discussions in the following section are centred on the definition and properties of the type of (super-)efficiency measures in (7), (10), and (12), regardless of the specific chosen centralized DEA programs behind.

4. Critical analysis, pitfalls and requirements

4.1. Basic data sets and computational results

Consider the example given in Table 1, which was introduced by Afsharian et al. (2017) and also used by Fang (2020). In this example, there are eight units with two inputs and with one output, which has the level of one for all units.

Table 1. Data of the example of 8 units

Unit	Input 1	Input 2	Output
1	6	1	1
2	2	2	1
3	1	5	1
4	4	4	1
5	4.5	3	1
6	3	4	1
7	5	4	1
8	3	5	1

The results of all measures outlined in the previous section are given in Table 2.

Table 2. Results for the example of 8 units

Unit	Decentralized (A & P 1993)	Centralized (Intermediate)	Centralized (Varmaz et al. 2013)	Centralized (Afsharian et al. 2017)	Centralized (Fang 2020)
	θ_p (θ_p^s)	θ_p^{int}	$\tilde{\theta}_p^{Va}$	$\tilde{\theta}_p^{Af}$	$\tilde{\theta}_p^{Fa}$
1	1.000 (2.000)	1.000	0.961	1.004	0.963
2	1.000 (1.611)	1.000	0.662	1.604	1.703
3	1.000 (2.000)	0.476	0.972	0.476	0.972
4	0.500 (0.500)	0.500	0.980	0.500	0.980
5	0.606 (0.606)	0.606	0.986	0.606	0.986
6	0.615 (0.615)	0.526	0.988	0.526	0.988
7	0.476 (0.476)	0.476	0.961	0.476	0.961
8	0.571 (0.571)	0.435	0.956	0.435	0.956

In each of the three outlined papers, the authors have also analyzed the performance of 16 branches of a German retail bank. The respective data set given in Table 3 was originally introduced by Varmaz et al. (2013).

Table 3. Input and output data of a German retail bank

Unit	PEX (input 1)	IEX (input 2)	IIN (output 1)	OIN (output 2)
1	1532.00	2769.00	11092.00	1231.00
2	998.00	1757.00	5529.00	778.00
3	853.00	1220.00	2384.00	464.00
4	180.00	378.00	632.00	133.00
5	584.00	876.00	1847.00	297.00
6	498.00	2080.00	2689.00	524.00
7	261.00	395.00	1358.00	203.00
8	609.00	883.00	2688.00	352.00
9	222.00	528.00	791.00	149.00
10	264.00	700.00	856.00	193.00
11	1078.00	1448.00	1873.00	611.00
12	222.00	503.00	770.00	217.00
13	258.00	412.00	520.00	138.00
14	696.00	1099.00	2836.00	443.00
15	176.00	361.00	477.00	104.00
16	236.00	301.00	724.00	159.00

The two inputs are personnel expenses (PEX) and expenses on interest payments (IEX), while the two outputs are interest income (IIN) and all other income (OIN). For a detailed description of these inputs and outputs, see Varmaz et al. (2013). The results generated by the three approaches are reproduced in Table 4.

Table 4. Efficiency scores of the branches of a retail bank

Branch	Decentralized (A & P 1993)	Centralized (Intermediate)	Centralized (Varmaz et al. 2013)	Centralized (Afsharian et al. 2017)	Centralized (Fang 2020)
	θ_p (θ_p^s)	θ_p^{int}	$\tilde{\theta}_p^{Va}$	$\tilde{\theta}_p^{Af}$	$\tilde{\theta}_p^{Fa}$
1	1.000 (N.E.*)	1.000	1.033	1.014	1.058
2	0.981 (0.981)	0.978	1.022	0.978	1.022
3	0.818 (0.818)	0.753	0.991	0.753	0.991
4	1.000 (1.075)	0.755	0.998	0.755	0.998
5	0.699 (0.699)	0.674	0.988	0.674	0.988
6	1.000 (1.242)	0.735	0.987	0.735	0.987
7	1.000 (1.182)	1.000	0.995	1.012	1.029
8	0.837 (0.837)	0.786	0.996	0.786	0.996
9	0.892 (0.892)	0.643	0.993	0.643	0.993
10	0.831 (0.831)	0.669	0.992	0.669	0.992
11	0.923 (0.923)	0.817	0.997	0.817	0.997
12	1.000 (1.176)	1.000	0.982	1.027	1.034
13	0.807 (0.807)	0.643	0.993	0.643	0.993
14	0.864 (0.864)	0.835	1.000	0.835	1.000
15	1.000 (1.037)	0.590	0.993	0.590	0.993
16	1.000 (1.262)	0.924	1.002	0.924	1.002

* This unit is not enveloped (for more details, see Varmaz et al. 2013 or Afsharian et al. 2017).

4.2. The issue of over-compensating operating units

As a starting point, compare the results of Fang's approach to those of the corresponding decentralized one (A & P: Andersen and Petersen (1993); the corresponding super-efficiency scores are given in parentheses). As can be seen in Table 2, e.g., U_7 is recognized as highly inefficient (0.476) by the decentralized approach. In contrast, this unit is identified as quite efficient (0.961) by Fang's approach. This result is also shared by Varmaz et al.'s approach, producing the same efficiency score of 0.961.

Let us reconsider the decentralized approach outlined in the previous section in conjunction with the DEA program in (3). The program is run n times in a row to compute the efficiency score for each DMU_p ($p = 1,\ldots,n$). This indicates that the program maximizes the individual efficiency of each unit relative to the other units in the system. As reasoned in the literature, a mechanism of this kind allows the DMUs to appear in their best possible light (see, e.g., Doyle and Green 1994; Portela et al. 2003; Camanho and Dyson 2005). Such an important property has led to the variety of applications of conventional DEA programs where units, e.g., operate independently, pursue their own individual goals, and make decisions or take actions based on their individual requirements, capabilities and priorities (see Emrouznejad and Yang 2018).

The above way of measuring the efficiency may not, however, be a proper approach in situations where the objective is to maximize the efficiency of the whole system rather than to optimize the individual efficiency of each unit. In other words – as stated in the literature – conventional DEA programs are not suitable to put the whole system of units in its best possible light (for a detailed discussion see, e.g., Roll et al. 1991, Lozano and Villa 2004; Kao and Hung 2005; Ruiz and Sirvent 2016; Hatami-Marbini et al. 2015, Fang 2020; Afsharian 2021). For example, an individual unit may choose to maximize its own efficiency that may not be optimal for the organization as a whole in terms of resource use relative to outcomes. Hence, this unit may then be recognized as less efficient by a centralized approach than by a decentralized one.

Against this background, the three outlined systems of incentives apply at their core centralized DEA programs such as the one in (6). With such a structure, these three approaches aim at putting the whole system of units in its best possible light by switching from a decentralized system of incentives to a centralized one. In line of the above, it is expected that the centralized approaches of Varmaz et al. (2013), Afsharian et al. (2017), and Fang (2020) have generally been designed in such a way that they should not lead to over-compensating units that had been inefficient under the decentralized approach, unless these units are identified as being, e.g., "jointly super-efficient" in the manner discussed in Section 4.4. More specifically:

Requirement #1: *The (super-)efficiency score of a unit determined by a centralized approach should not generally exceed its (super-)efficiency score determined by the decentralized approach.*

However, this is not the case in Fang's or in Varmaz et al.'s approach. U_7, e.g., which would have received a significant penalty in the decentralized framework (where its degree of autonomy is the highest), receives a very generous assessment under central management by these two approaches. This inconsistency can also be observed for almost all other units in Table 2. In contrast, Afsharian et al.'s approach has been shown to always produce results consistent with Requirement 1 (see also Theorem 2 in Afsharian et al. 2017). As will be shown later, Afsharian et al.'s approach is also unable to provide a comprehensive solution when it comes to exceptional cases like the one with units that are identified as "jointly super-efficient" (see the discussions in Section 4.4.).

Let us emphasize here that Requirement 1 does not rule out the opposite case: An efficient unit under decentralized management may not necessary be efficient in the respective centralized framework. For example, in Afsharian et al.'s approach, U_1 is efficient in the decentralized setting, but it is not efficient in the centralized setting.

The detected counter-intuitive results in Varmaz et al.'s and Fang's approaches are not limited to the theoretical example depicted in Table 2. Consider branch 2 in Table 4. It is recognized as inefficient (0.981) by the decentralized approach. In contrast, both Varmaz et al.'s approach and Fang's approach determine a super-efficiency score of 1.022 for this branch, suggesting that it should even be rewarded. For example, Fang (2020, p. 159) states that

> "… although branch 2 is technically[5] inefficient, it positively contributes to the system's overall performance."

The author argues further:

> "In fact, if we use the multipliers in determining the overall efficiency by model (3)[6] to calculate the efficiency score of branch 2, its efficiency score[7] equals 0.978, which outweighs the system's overall efficiency (0.835) and thus has a positive effect on the system's overall performance."

This means that if a unit shows a higher intermediate efficiency score compared to the system's overall efficiency (i.e., $\theta_p^{int} \geq Eff$), the unit should receive a reward in the system of incentives. Considering the results in Table 4, this is fulfilled in this particular case: for those branches that have an intermediate efficiency score (see the third column) greater that the overall efficiency (computed as $Eff = 0.835$), Fang's approach assigns a super-efficiency score (see branches 1, 2, 7, 12 and 16 in the last column in Table 4).

---

[5] Determined by the decentralized approach.

[6] Our program in (6).

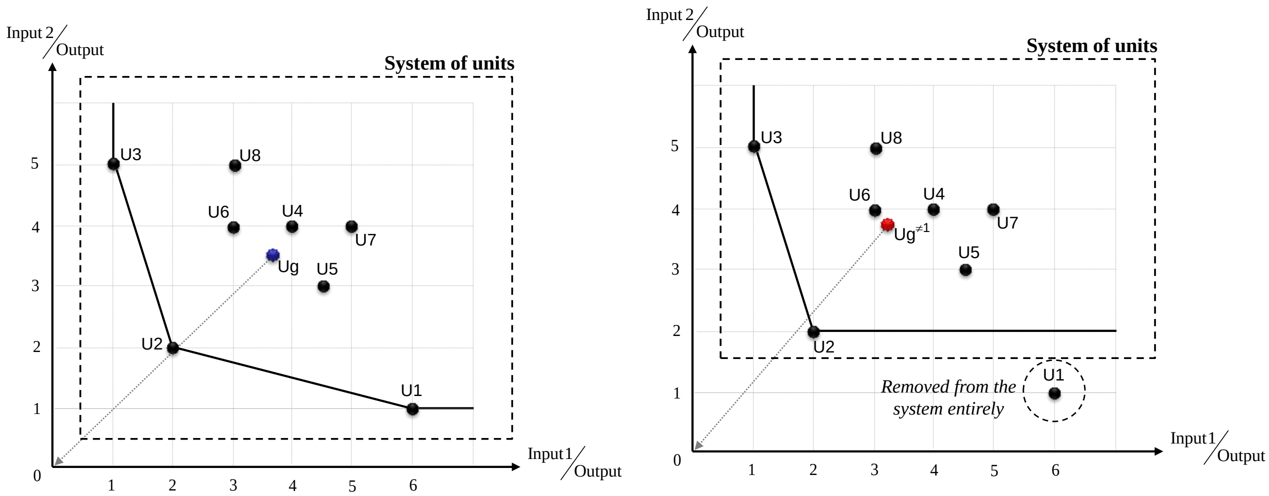
[7] With our notations, its intermediate efficiency score, i.e., $\theta_2^{int} = 0.978$.

With this argument from Fang (2020) at hand, let us turn to the example of eight units and the corresponding results in Table 2, and consider U_1. A representation of the system of units is given in Figure 1(a). According to the discussion in Section 3, the overall efficiency of the whole system is measured by the efficiency of the grand unit that has the average value of inputs and outputs computed across all units in the system, i.e., see the program in (6). In Figure 1(a), this grand unit is indicated by U_g.

The system's overall efficiency is computed as $Eff = 0.569$ where the benchmark of U_g are U_1 and U_2. Now, with the same multipliers (associated with the line between U_1 and U_2) that are used for determining the overall efficiency, U_1 is fully efficient (see the intermediate efficiency score in the third column: θ_1^{int} = 100%). This shows that the efficiency of this unit clearly outweighs the system's overall efficiency, i.e., $\theta_1^{int} \geq Eff$. With the above line of reasoning given by Fang (2020), this unit has a positive impact on the system's overall performance and should thus receive a reward, i.e., this should be reflected in a super-efficiency score by Fang's approach higher than 1. However, this is not the case: Fang's approach captures a negative impact of $\tilde{\theta}_1^{Fa} = 0.963$ (see the last column in Table 2), implying that this unit is not even recognized as efficient, thus receiving a penalty! This is clearly inconsistent with the author's arguments themselves. This problem is also shared with Varmaz et al.'s approach, which assigns a score of 0.931 to U_1 (see the fourth column of Table 2).

As outlined in the previous section, in order to measure the super-efficiency of each unit under central management, Varmaz et al.'s approach compares the efficiency of the whole system with the efficiency score of the system where the unit under evaluation is excluded entirely both from the potential referent units and from the computation of the grand unit. The key pitfall here is that such a comparison is not defined in a stable manner, i.e., the numerator and the denominator of $\tilde{\theta}_p^{Va} = Eff \,/\, Eff_{\neq p}^{\neq p}$ refer to "different systems of relative efficiency", where an altering reference point (in the form of a grand unit) is used once with and once without DMU_p. These two systems of relative efficiency are represented graphically in Figure 1 for unit 1, c.f., Figure 1(a) and 1(b) to observe that the position of the grand unit (U_g and $U_g^{\neq 1}$) alters from one system to the other.

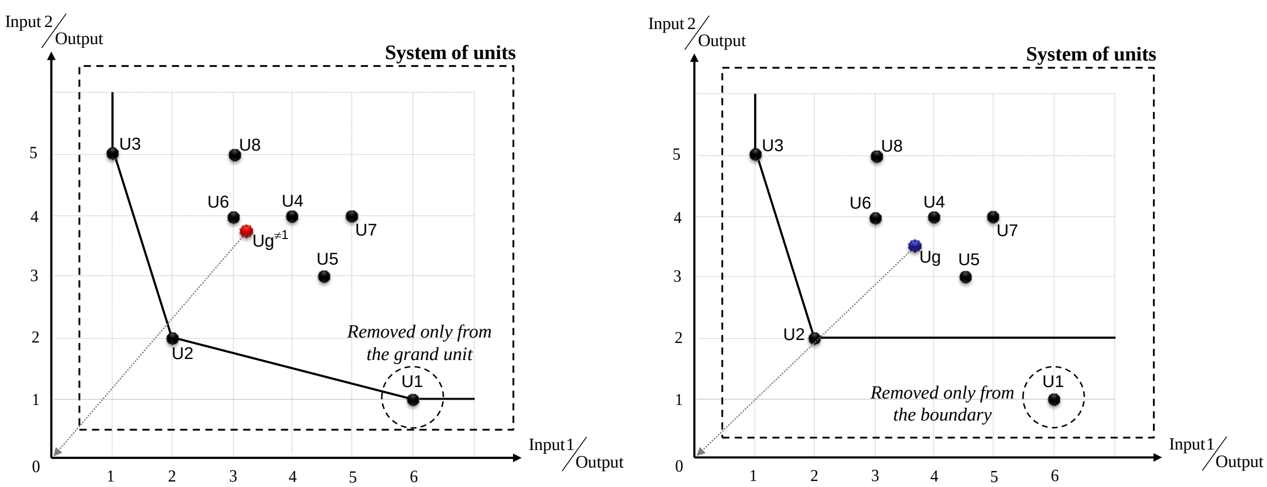
Figure 1. Representations of the system of 8 units within Varmaz et al.'s approach



(a) (b)

Fang (2020) suggests the super-efficiency measure $\tilde{\theta}_p^{Fa} = Eff_{\neq p} / Eff^{\neq p}$. Here, the unit under evaluation is excluded once from the candidate boundary units in the numerator (where the grand unit remains unchanged) and once from the grand unit in the denominator (where the boundary remains intact). See Figure 2 in which the two systems of relative efficiency for U_1 are represented graphically.

Figure 2. Representations of the system of 8 units within Fang's approach

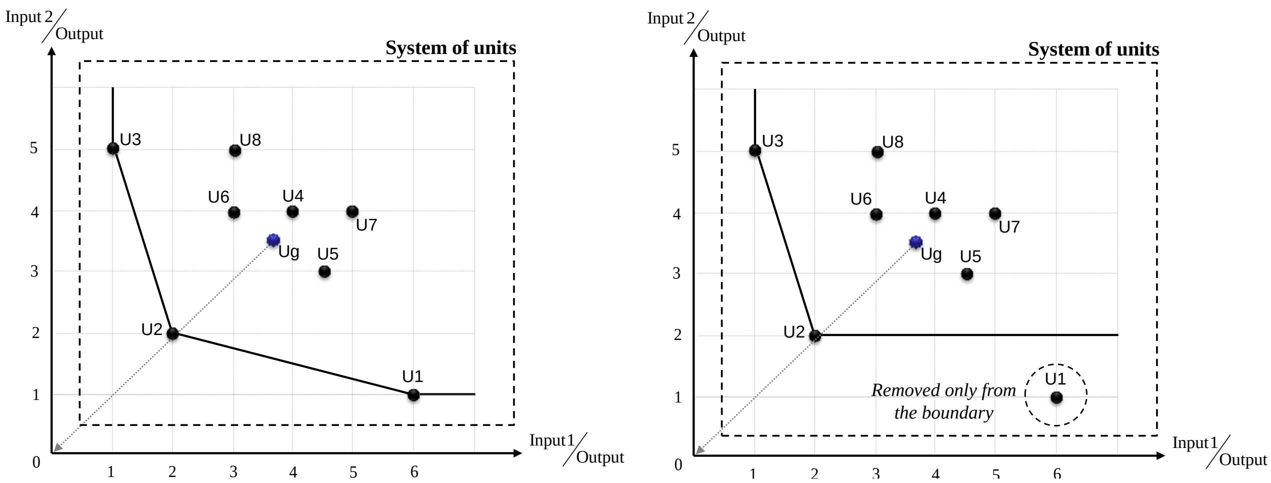


(a) (b)

As can be observed, Fang's approach suffers from the same issue as Varmaz et al.'s approach because of the use of an altering grand unit (U_g and $U_g^{\neq 1}$) in the corresponding ratio, i.e., the numerator and the denominator refer to different systems of relative efficiency.

As exemplified above, such a drawback produces counter-intuitive results, incompatible with incentivizing units to improve their performance. We observe, however, that Afsharian et al.'s approach is immune to this particular criticism as in the proposed measure of super-efficiency – $\tilde{\theta}_p^{Af} = \left(Eff_{\neq p} / Eff\right) \times \theta_p^{int}$ – the unit under evaluation is excluded from candidate boundary units only, c.f., Figure 3(a) and 3(b). This provides a constant reference point (U_g) – and accordingly constant benchmarks – for capturing the changes in the efficient boundary, i.e., the numerator and the denominator in $Eff_{\neq p} / Eff$ refer to the same systems of relative efficiency.

Figure 3. Representations of the system of 8 units within Afsharian et al.'s approach



(a) (b)

Furthermore, unlike in the approaches of Varmaz et al. (2013) and Afsharian et al. (2017), it is not clear which units are benchmarks for the system of units as a whole in the approach of Fang (2020). In both Varmaz et al.'s and Afsharian et al.'s approach, the benchmark units (captured while the efficiency of the whole system of units is measured by the centralized DEA program in (6)) are U_1 and U_2, as depicted in Figure 1(a) and Figure 3(a). However, the application of Fang's approach may lead to altering benchmarks for the system of units as a whole, as can be seen in Figure 2(a). In particular, U_3, which is identified as benchmark (i.e., a good performing unit from the perspective of Fang's approach), receives a penalty, which is not compatible with the grounded concept of incentivization discussed in Section 3: This would discourage U_3 to perform still better by not further revealing information that push out the frontier, which could have in turn lead to (more) cost-efficient behaviour by the units in the future (see also the discussions in the next section).

4.3. The issue of the ratchet effect

Let us refer to Varmaz et al. (2013, p. 113) who stated that

> "… the performance estimator has to be able to take values above 1. If this requirement is not met, agents would only receive negative incentives, i.e., punishments for performing worse than best practice. Consequently, they would only try to perform as good as best practice, but would have no incentives for further improvements", i.e., the ratchet effect would occur.

We agree with Varmaz et al. (2013) that this is of particular importance in practice because the central management wishes to push out progressively to more productive positions the frontier by setting appropriate incentives (i.e., more demanding targets) from one evaluation period to the next (see Bottasso and Conti 2009). Hence, it is essential in systems of incentives that an approach is designed in a way that the results can encourage all the units – including the already efficient ones – to operate as efficiently as possible. In particular, the already efficient units should be rewarded to reveal further efficient cost levels, eliminating the ratchet effect (see, e.g., Agrell et al. 2005).

A closer look at the results of Varmaz et al.'s approach in Table 2 reveals that there is no unit with a positive impact on the overall efficiency of the whole system, i.e., all scores are less than one. This means that all units are considered inefficient, being penalized within this system of incentives. First of all, this is inconsistent with the definition of "relative efficiency" itself. Furthermore, these results contradict the concept of "super-efficiency" and the above statement of the author by which high-powered incentives should be given to units with exceptional performance. As an example, consider U_2 in Figure 1(a). This unit is identified efficient by the centralized program in (6), i.e., its intermediate efficiency score $\theta_2^{int} = 100\%$. Hence, it serves in this case as the benchmark for U_g, meaning that other units in the system should emulate this unit in order for the efficiency of the organization to improve. Nevertheless, this unit – which also interestingly utilizes the resources in a balanced way – is severely punished by Varmaz et al.'s approach. Analysing the definition of the measure $\tilde{\theta}_p^{Va} = Eff \,/\, Eff_{\neq p}^{\neq p}$ in Varmaz et al.'s approach can clarify this issue:

We recall that Eff in the numerator is the efficiency of the whole system of units (i.e., overall efficiency) measured by the centralized DEA program in (6). Let us assume that with the same multipliers that are used for determining the overall efficiency, a unit (like U_2 in our graphical example) is efficient, i.e., by the notation in (9), $\theta_p^{int} = 100\%$. Removing such a unit entirely from the data set may not necessarily result in a value for the denominator $Eff_{\neq p}^{\neq p}$ to be less than or equal Eff. This means that $\tilde{\theta}_p^{Va} = Eff \,/\, Eff_{\neq p}^{\neq p}$ can ultimately be less than one, which is counter to expectation. For instance, in our graphical example, Eff = 0.569, U_2 is efficient ($\theta_2^{int} = 100\%$) and $Eff_{\neq 2}^{\neq 2}$ = 1.510. This results in $\tilde{\theta}_2^{Va} = 0.662$, implying that this efficient unit is severely punished within this system of incentives. The same also happens for the branches 7 and 12 in the case of banking. Although they are efficient, as their intermediate efficiency scores reveal in the third column of Table 4, they are ultimately recognized as inefficient by Varmaz et al.'s approach, being penalized within the system of incentives.

To formalize this critical issue, we define the following requirement:

Requirement #2: *A well-defined efficiency measure under central management should be consistent with the concept of "relative efficiency" and also should offer incentives to those units recognized as efficient to act in the best interest of the entire organization.*

To see if Fang's approach fulfils this requirement, consider again the definition of the efficiency measure in this approach: $\tilde{\theta}_p^{Fa} = Eff_{\neq p} / Eff^{\neq p}$. Let us assume that in a particular system of units, a DMU$_p$ is already efficient by the centralized program in (6), i.e., $\theta_p^{int} = 100\%$. Hence, removing this unit from the boundary unit must result in the numerator to be $Eff_{\neq p} \geq Eff$. Let us suppose that removing DMU$_p$ from the grand unit leads to an alternate grand unit so that $Eff^{\neq p} > Eff$. Now, in situations where $Eff^{\neq p} > Eff_{\neq p}$, we will have $\tilde{\theta}_p^{Fa} < 1$, even if DMU$_p$ is already efficient by the centralized program in (6). This is the case for U$_1$ in our graphical example. For this efficient unit ($\theta_1^{int} = 100\%$), $Eff^{\neq 1} = 0.592$ and $Eff_{\neq 1} = 0.571$, implying that $Eff^{\neq 1} > Eff_{\neq 1}$. This results in a value of ultimate efficiency $\tilde{\theta}_1^{Fa} = Eff_{\neq 1} / Eff^{\neq 1}$ less than one (i.e., $\tilde{\theta}^{Fa} = 0.963$), which is counter-intuitive.

An investigation of this issue in Afsharian et al.'s approach shows that this requirement is always fulfilled. Consider again the definition of the efficiency measure $\tilde{\theta}_p^{Af} = \left(Eff_{\neq p} / Eff \right) \times \theta_p^{int}$ in this approach. Let us assume that with the same multipliers that are used for determining the overall efficiency, a DMU$_p$ is efficient, i.e., $\theta_p^{int} = 100\%$. Two scenarios may now arise:

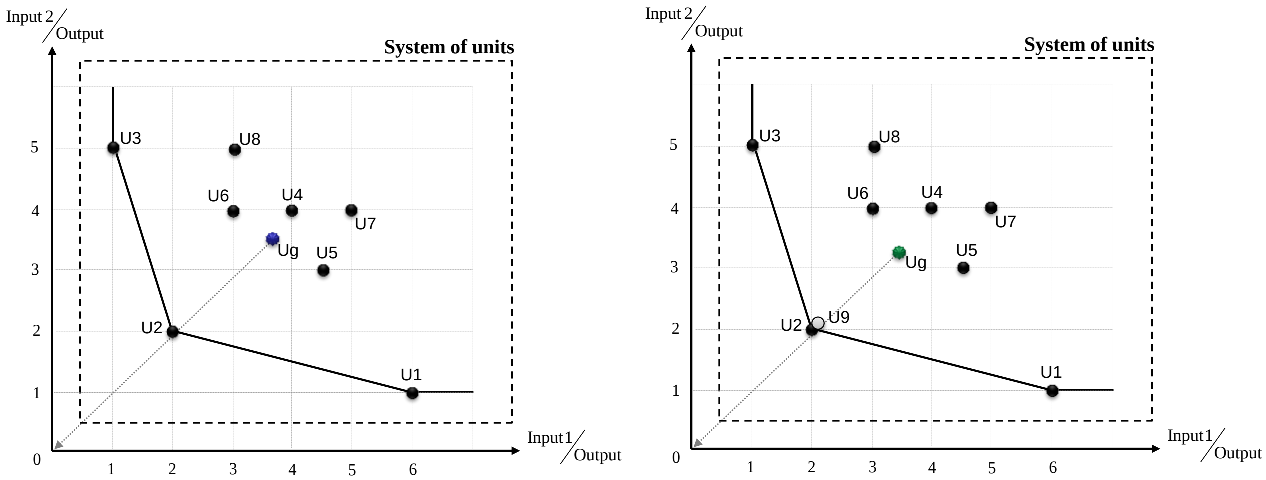
- The exclusion of this unit from the boundary affects the assessed performance of the whole system, representing by the grand unit. In this case, as $Eff_{\neq p}$ will always be greater than Eff, then $\tilde{\theta}_p^{Af} > 100\%$.
- The exclusion of DMU$_p$ does not affect the assessed performance of the whole system. This happens in rare circumstances that removing an efficient unit from the boundary does not affect the boundary (e.g., because the unit concerned is a linear combination of other efficient units). In this case, $Eff_{\neq p}$ equals Eff, which results in $\tilde{\theta}_p^{Af} = 100\%$.

4.4. The issue of masked performance

The above-outlined systems of incentives apply types of conventional super-efficiency in the sense that the impact of the exclusion of "one unit at a time" from the efficient frontier is captured. As a consequence, these approaches may not identify properly the impact of certain units that have a very similar performance (they mask the performance of each other even if jointly they are substantially different from other units in the system).

In order to illustrate the issue, consider once again our graphical example. Let us now assume that the initial set of units was not the 8 depicted in Figure 4(a) but rather the 9 depicted in Figure 4(b). Note that existing units U_2 and U_9 have input-output levels which are very similar in levels and mix, and in addition, they both have significantly better performance than the rest of the units employing a similar mix of inputs and outputs.

Figure 4. Representation of the system of units without and with an additional unit U_9



(a) (b)

The new results obtained by applying the three approaches are reported in Table 5.

Table 5. Results for the example of 9 units

Unit	Decentralized (A & P 1993)	Centralized (Intermediate)	Centralized (Varmaz et al. 2013)	Centralized (Afsharian et al. 2017)	Centralized (Fang 2020)
	θ_p (θ_p^s)	θ_p^{int}	$\tilde{\theta}_p^{Va}$	$\tilde{\theta}_p^{Af}$	$\tilde{\theta}_p^{Fa}$
1	1.000 (2.000)	1.000	0.959	1.003	0.962
2	1.000 (1.100)	1.000	0.956	1.099	1.155
3	1.000 (2.000)	0.476	0.969	0.476	0.969
4	0.500 (0.500)	0.500	0.976	0.500	0.976
5	0.606 (0.606)	0.606	0.982	0.606	0.982
6	0.615 (0.615)	0.526	0.984	0.526	0.984
7	0.476 (0.476)	0.476	0.959	0.476	0.959
8	0.571 (0.571)	0.435	0.954	0.435	0.954
9	0.909 (0.909)	0.909	1.043	0.909	1.043

In Fang's approach, the efficiency of U_2 has decreased significantly from its previous value $\tilde{\theta}_2^{Fa} = 1.703$ – where U_9 was not within the original set of units – to the new value of $\tilde{\theta}_2^{Fa} = 1.155$. Note that this approach excludes each unit at a time once from the candidate boundary units and once from the grand unit. Excluding U_2 from the boundary units leads to a very similar shape of the boundary, resulting in $Eff_{\neq 2} = 0.653$, c.f. the system's overall efficiency computed as $Eff = 0.594$. The position of the grand unit does not alter either when U_2 is removed from the grand unit, resulting in $Eff^{\neq 2} = 0.565$, c.f. $Eff = 0.594$. This produces an efficiency value of 1.155 within the formula $\tilde{\theta}_2^{Fa} = Eff_{\neq 2} / Eff^{\neq 2}$, which is considerably different from its previous value 1.703. With respect to U_9, its efficiency is computed as $\tilde{\theta}_9^{Fa} = 1.043$, suggesting this unit be rewarded. This is a good result in this particular example, although the value is substantially less than 1.703, which could have been obtained if U_2 was not in the analysis (i.e., if we did not have units U2 and U9 masking one another when their super-efficiencies are computed). However, it comes at the costs of not being able to provide consistent results in the context of other requirements in the previous section.

As Varmaz et al.'s approach removes each unit at a time entirely from the data set, the results are not even plausible in this case: As in Fang's approach, the efficiency of U_9 is $\tilde{\theta}_9^{Fa} = 1.043$ in Varmaz et al.'s approach. However, the efficiency score of U_2 has increased significantly from its previous value $\tilde{\theta}_2^{Va} = 0.662$ to the new value of $\tilde{\theta}_2^{Va} = 0.956$!

Afsharian et al.'s approach also suffers from the same issue. In this approach, each unit is excluded at a time from the boundary units only. Hence, on the one hand, the efficiency of U_2 has reduced from $\tilde{\theta}_2^{Af} = 1.604$ to the new value of $\tilde{\theta}_2^{Af} = 1.099$, because of a very similar shape of the boundary after excluding U_2 from the boundary units, c.f. $Eff_{\neq 2} = 0.653$ and $Eff = 0.594$. On the other hand, as U_9 is inefficient (i.e., θ_9^{int} = 0.909), the exclusion of this unit does not affect the performance of the whole system, i.e., $Eff_{\neq p} = Eff$ = 0.594. Accordingly, the efficiency of U_9 is represented by a value of $\tilde{\theta}_p^{Af} = 0.909$. This is considerably less than 1.604, which could have been approximately obtained if U_2 was not in the analysis.

In this example, this issue has happened because units 2 and 9 mask the outstanding performance of each other. This shows that the above approaches may not appropriately identify units that are near neighbours in inputs and outputs (e.g., two units that "mask" each other) as "significantly" super-efficient, even if they jointly have a substantial impact on the system's overall efficiency. Calling attention to this potential problem, we emphasize a third requirement:

Requirement #3: *An appropriate system of incentives under central management should be equipped with an instrument to identify properly the impact of certain units that are near neighbours in inputs and outputs.*

The approaches by Varmaz et al. (2013) and Fang (2020) are silent to this issue. Afsharian et al. (2017) acknowledge its existence in their proposed approach and suggest a complementary analysis for overcoming this potential problem. Their simple technique is an adaptation of the sensitivity-based procedure introduced by Thanassoulis (1999). In a nutshell, within this technique, units with an intermediate efficiency $\theta_p^{int} \geq (100 - r)\%$ are collected, where 100-*r* is a user-specified efficiency level which is close enough to

100% to be deemed as 100%. Through a particular process, these units are excluded all together from the boundary to capture their impact (see the details in Section 5 of Afsharian et al. 2017).

For example, applying their approach with r = 10, the new efficiencies of U_1, U_2 and U_9 (those units whose previous efficiency scores were greater than or equal 90%) are 1.768, 1.768 and 1.759, respectively. As can be seen, similar to U_1 and U_2, U_9 is also rewarded. This is done to encourage the unit to perform better still, which will reduce the asymmetry of information about cost-efficient behaviour by the units in future. Note that we cannot generally argue that U_9 (which uses the resources in a balanced way in this particular graphical example) should be given a higher reward than U_1 because these two units operate in different mixes of inputs. Offering reward and its magnitude is associated to the impact each unit has on the grand unit and the way the unit assessment can push out the frontier. Nevertheless, one may add complementary criteria or develop further the methods to also reflect such characteristics in the reward given to a unit (see also Section 5).

While Afsharian et al.'s approach deals with the issue of masked performance, it is unable to provide a comprehensive solution. From the example, in can be seen that the previous difference between U_1 and U_2 captured by this approach (reflecting a higher contribution of U_2 as a benchmark for the grand unit compared to U_1) disappeared. Now, both U_1 and U_2 are rewarded equally, which can be considered naïve per se. Hence, we believe that an enhancement of the approach (the formula in (10)) is required in a way that it could solve the issue of masked performance while at the same time the other properties remain fulfilled.

5. Conclusion and outlook on future research

We have given a critical overview of the approaches of Varmaz et al. (2013), Afsharian et al. (2017) and Fang (2020) designed for incentivizing operating units in multi-unit organizations to reveal progressively more efficient operating practices. Within these approaches, while inefficient units are encouraged to make efficiency savings, units which are identified to be efficient are incentivized by some reward that should be consistent with the level of their impact on the overall performance of the organization. We have underlined essential requirements necessary in designing such a well-defined system of incentives under central management, in particular dealing with the issue of appropriately compensating units, the issue of avoiding the ratchet effect and the issue of handling units masking each other's performance.

The paper aims at enhancing the awareness of the discussed pitfalls and requirements amongst researchers and particularly at encouraging further research in this field. The requirements we have introduced here lay the foundations of a framework for developing a system of incentivizing units to operate in a manner consistent with improving the overall performance of an organization where central management does not manage directly the operations of units. Clearly, further research is needed in developing systems of this type. We believe that the perspective and the techniques discussed in this paper create respective opportunities in the context of DEA-based system of incentives under central management:

- In the domain of incentivization, it is crucial to show that the results of an approach are incentive-compatible in the sense that the units under evaluation ultimately prefer a cost strategy without slack. For the case under decentralized management, there exists supporting research (see, e.g., Bogetoft 1997; Agrell et al. 2005; Bogetoft 2013). In the context of incentivization under central management, although some discussions have already been given in Section 2, a more comprehensive theoretical foundation is required. In particular, the methods reviewed here (Fang 2020, Afsharian et al. 2017) and Varmaz et al. 2013) offer pragmatic approaches to incentivizing units to improve efficiency in a centralized context. The theory by Bogetoft (1997), from which they draw, addresses the issue of utility of the unit being compensated as it is affected by the level of compensation it is offered. The level of compensation is the instrument by which that utility is affected, which in turn incentivizes the unit to reveal more efficient practices. None of the three approaches discussed here has transferred that theory into the centralized approach. It is noteworthy that the utility addressed in the centralized context may be not only that of the unit being compensated but also that of the central organization. The latter suffers a more direct impact, including to its profits, from the operations of the constituent units than is the case in economic regulation of independent entities. The respective transfer of the theory underpinning optimal compensation from the decentralized to the centralized context remains an area for further research.

- Strong monotonicity in efficiency under decentralized management generally means that any increase in input and decrease in output should decrease the efficiency of the unit under assessment, when other units' input/output mixes and levels are unchanged (see, e.g., Bogetoft and Hougaard 1999; An et al. 2022). As the fulfilment of this requirement is a desirable property both at unit and grand central unit level, it should be investigated which modification of the strong monotonicity is needed in the centralized context.

- In the domain of incentive regulation, several goals must be balanced (Antle and Bogetoft 2019). It is worth considering – in a similar way presented in, e.g., Bogetoft and Eskesen (2022) – to develop further an approach by which the effect of the mix of inputs/outputs can be reflected more appropriately in the design of incentives under central management. In the same line, an appropriate modification of the approach proposed by Afsharian et al. (2019) – which cluster the units based on their mix of inputs – could be useful.

- We have shown that the sensitivity-based procedure to deal with the issue of masked performance addressed in Section 4.4 is not optimal. Following the discussions in this section, the mixed-integer program-based method recently proposed by Afsharian and Bogetoft (2020) is noteworthy. This technique can identify – in a controlled manner – the subset of *k* units, which when removed from the efficient frontier, yields the greatest impact on the overall efficiency of the whole system of *n* units.[8] Future research on this topic seems to be fruitful.

Acknowledgment

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) in the context of the research fund AH 90/5-3.

References

Afsharian, M. (2020). A metafrontier-based yardstick competition mechanism for incentivising units in centrally managed multi-group organisations. Annals of Operations Research, 288(2), 681-700.

Afsharian, M. (2021). A frontier-based facility location problem with a centralised view of measuring the performance of the network. Journal of the Operational Research Society, 72(5), 1058-1074.

Afsharian, M. (2022). Prescriptive analytics in benchmarking, incentive regulation and contract design, Habilitation, Univers.-Bibl. Braunschweig.

Afsharian, M., Ahn, H. (2017). Multi-period productivity measurement under centralized management with an empirical illustration to German saving banks. OR Spectrum, 39(3), 881-911.

Afsharian, M., Ahn, H., Harms, S. G. (2021). A review of DEA approaches applying a common set of weights: The perspective of centralized management. European Journal of Operational Research, 294(1), 3-15.

Afsharian, M., Ahn, H., Kamali, S. (2022). Performance analytics in incentive regulation: A literature review of DEA publications. Decision Analytics Journal, 100079. (https://doi.org/10.1016/j.dajour.2022.100079)

Afsharian, M., Ahn, H., Thanassoulis, E. (2017). A DEA-based incentives system for centrally managed multi-unit organisations. European Journal of Operational Research, 259(2), 587-598.

Afsharian, M., Ahn, H., Thanassoulis, E. (2019). A frontier-based system of incentives for units in organisations with varying degrees of decentralisation. European Journal of Operational Research, 275(1), 224-237.

Afsharian, M., Bogetoft, P. (2020). Identifying production units with outstanding performance. European Journal of Operational Research, 287(3), 1191-1194.

Agrell, P. J., Bogetoft, P. (2017). Regulatory benchmarking: Models, analyses and applications. Data Envelopment Analysis Journal, 3(1-2), 49-91.

[8] As the authors illustrate their approach with the same banking data set given in Table 3, an example could be useful: Let us assume that we seek to find the subset of *k* outstanding branches that have the greatest impact on the overall efficiency of the whole system of 16 branches. Applying their mixed-integer program, in case *k*=1, branch 12 is recognized as outstanding. This result is not surprising as this branch shows the highest super-efficiency score of 1.027 within Afsharian et al.'s approach. However, in case of, e.g., *k*=3, the identified branches are 1, 2 and 7. This means that this subset of three branches has an impact, which is the largest compared to any other subset of three branches that could have been selected. Interesting to note that – in Afsharian et al.'s approach – while branches 1 and 7 are already recognized as super-efficient, branch 7 has an efficiency score of 0.978. Nevertheless, the mixed-integer program also suggests branch 7 as outstanding. For more details, see Section 4 in Afsharian and Bogetoft (2020).

Agrell, P. J., Bogetoft, P. (2018). Theory, techniques, and applications of regulatory benchmarking and productivity analysis. The Oxford handbook of productivity analysis. Oxford University Press, Oxford.

Agrell, P. J., Bogetoft, P., Tind, J. (2002). Incentive plans for productive efficiency, innovation and learning. International Journal of Production Economics, 78(1), 1-11.

Agrell, P. J., Bogetoft, P., Tind, J. (2005). DEA and dynamic yardstick competition in Scandinavian electricity distribution. Journal of Productivity Analysis, 23(2), 173-201.

An, Q., Tao, X., Xiong, B., Chen, X. (2022). Frontier-based incentive mechanisms for allocating common revenues or fixed costs. European Journal of Operational Research, 302(1), 294-308.

Andersen, P., Petersen, N. C. (1993). A procedure for ranking efficient units in data envelopment analysis. Management Science, 39(10), 1261-1264.

Anthony, R. N., Govindarajan, V., Hartmann, F. G. H., Kraus, K., Nilsson, G. (2014). Management control systems. McGraw-Hill Higher Education, Berkshire.

Antle, R., Bogetoft, P. (2019). Mix stickiness under asymmetric cost information. Management Science, 65(6), 2787-2812.

Asmild, M., Paradi, J. C., Pastor, J. T. (2009). Centralized resource allocation BCC models. Omega, 37(1), 40-49.

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.

Birnberg, J. G., Snodgrass, C. (1988). Culture and control: A field study. Accounting, Organizations and Society, 13(5), 447-464.

Bogetoft, P. (1997). DEA-based yardstick competition: The optimality of best practice regulation. Annals of Operations Research, 73, 277–298.

Bogetoft, P. (2000). DEA and activity planning under asymmetric information. Journal of Productivity Analysis, 13(1), 7-48.

Bogetoft, P. (2013). Performance benchmarking: Measuring and managing performance. Springer, New York.

Bogetoft, P., Eskesen, A. (2022). Balancing information rents and service differentiation in utility regulation. Energy Economics, 105808. (https://doi.org/10.1016/j.eneco.2022.105808)

Bogetoft, P., Hougaard, J. L. (1999). Efficiency evaluations based on potential (non-proportional) improvements. Journal of Productivity Analysis, 12(3), 233-247.

Bonner, S. E., Sprinkle, G. B. (2002). The effects of monetary incentives on effort and task performance: Theories, evidence, and a framework for research. Accounting, Organizations and Society, 27(4-5), 303-345.

Bottasso, A., Conti, M. (2009). Price cap regulation and the ratchet effect: A generalized index approach. Journal of Productivity Analysis, 32(3), 191-201.

Camanho, A. S., Dyson, R. G. (2005). Cost efficiency measurement with price uncertainty: A DEA application to bank branch assessments. European Journal of Operational Research, 161(2), 432-446.

Dai, Q., Li, Y., Lei, X., Wu, D. (2021). A DEA-based incentive approach for allocating common revenues or fixed costs. European Journal of Operational Research, 292(2), 675-686.

Davtalab-Olyaie, M., Mahmudi-Baram, H., Asgharian, M. (2021). Incentivizing units in centralized systems: A slacks-based approach. Journal of the Operational Research Society (https://doi.org/10.1080/01605682.2021.1932620).

Doyle, J., Green, R. (1994). Efficiency and cross-efficiency in DEA: Derivations, meanings and uses. Journal of the Operational Research Society, 45(5), 567-578.

Emrouznejad, A., Yang, G. L. (2018). A survey and analysis of the first 40 years of scholarly literature in DEA: 1978–2016. Socio-Economic Planning Sciences, 61, 4-8.

Fang, L. (2020). An incentive approach based on data envelopment analysis for intra-organization yardstick competition. Journal of the Operational Research Society, 71(1), 153-160.

Flamholtz, E. G., Das, T. K., Tsui, A. S. (1985). Toward an integrative framework of organizational control. Accounting, Organizations and Society, 10(1), 35-50.

Green, S. G., Welsh, M. A. (1988). Cybernetics and dependence: Reframing the control concept. Academy of Management Review, 13(2), 287-301.

Hatami-Marbini, A., Tavana, M., Agrell, P. J., Lotfi, F. H., Beigi, Z. G. (2015). A common-weights DEA model for centralized resource reduction and target setting. Computers & Industrial Engineering, 79, 195-203.

Kao, C., Hung, H. T. (2005). Data Envelopment Analysis with common weights: The compromise solution approach. Journal of the Operational Research Society, 56(10), 1196-1203.

Lozano, S., Villa, G. (2004). Centralized resource allocation using data envelopment analysis. Journal of Productivity Analysis, 22(1-2), 143–161.

Malmi, T., Brown, D. A. (2008). Management control systems as a package: Opportunities, challenges and research directions. Management Accounting Research, 19(4), 287-300.

Mar-Molinero, C., Prior, D., Segovia, M. M., Portillo, F. (2014). On centralized resource utilization and its reallocation by using DEA. Annals of Operations Research, 221(1), 273-283.

Pernot, E., Roodhooft, F. (2014). The impact of inter-organizational management control systems on performance: A retrospective case study of an automotive supplier relationship. International Journal of Production Economics, 158(C), 156-170.

Portela, M. C. A. S., Borges, P. C., Thanassoulis, E. (2003). Finding closest targets in non-oriented DEA models: The case of convex and non-convex technologies. Journal of Productivity Analysis, 19(2-3), 251-269.

Roll, Y., Cook, W. D., Golany, B. (1991). Controlling factor weights in data envelopment analysis. IIE Transactions, 23(1) 2-9.

Ruiz, J. L., Sirvent, I. (2016). Common benchmarking and ranking of units with DEA. Omega, 65, 1-9.

Shleifer, A. (1985). A theory of yardstick competition. The RAND Journal of Economics, 16(3), 319-327.

Simons, R. (1987). Accounting control systems and business strategy: An empirical analysis. Accounting, Organizations and Society, 12(4), 357-374.

Thanassoulis, E. (1999). Setting achievement targets for school children. Education Economics, 7(2), 101-119.

Thanassoulis, E. (2001). Introduction to the theory and application of data envelopment analysis. Springer, New York.

Thanassoulis, E., Blake, C., Parthasarathy, S. (2022). Implicit and explicit incentives within economic regulation. Decision Analytics Journal, https://doi.org/10.1016/j.dajour.2022.100099.

Varmaz, A., Varwig, A., Poddig, T. (2013). Centralized resource planning and yardstick competition. Omega, 41(1), 112-118.

Ylvinger, S. (2000). Industry performance and structural efficiency measures: Solutions to problems in firm models. European Journal of Operational Research, 121(1), 164-174.