A DEA-based Incentives System for Centrally Managed Multi-Unit Organisations

Mohsen Afsharian[1a], Heinz Ahn[a], Emmanuel Thanassoulis[b]

[a] *Department of Business Sciences, Technische Universität Braunschweig, Fallersleber-Tor-Wall 23, 38100 Braunschweig, Germany*
[b] *Operations and Information Management Group, Aston Business School, Aston University, Birmingham B4 7ET, United Kingdom*

Abstract. In multi-unit organisations such as a bank and its branches or a national body delivering publicly funded health or education services through local operating units, the need arises to incentivize the units to operate efficiently. In such instances, it is generally accepted that units found to be inefficient can be encouraged to make efficiency savings. However, units which are found to be efficient need to be incentivized in a different manner. It has been suggested that efficient units could be incentivized by some reward compatible with the level to which their attainment exceeds that of the best of the rest, normally referred to as "super-efficiency". A recent approach to this issue (Varmaz et. al. 2013) has used Data Envelopment Analysis (DEA) models to measure the super-efficiency of the whole system of operating units with and without the involvement of each unit in turn in order to provide incentives. We identify shortcomings in this approach and use it as a starting point to develop a new DEA-based system for incentivizing operating units to operate efficiently for the benefit of the aggregate system of units. Data from a small German retail bank is used to illustrate our method.

Keywords: Data Envelopment Analysis (DEA); Centralised Management; Incentive Regulation; Banking

---

[1] Corresponding author. Email: m.afsharian@tu-bs.de; Phone: +495313913606; Fax: +495313918121

1. Introduction

In many instances in the public and the private sector, we encounter situations where a central body manages a large set of similar units delivering some services. Examples of such centrally managed multi-unit organisations are a bank managing its branches, a tax authority managing local tax offices, a supermarket chain managing its outlets etc. In such organizations, the operating units have various degrees of autonomy. The central management varies in the degree to which it controls the day to day operations of each unit, and typically there is a considerable degree of autonomy for each organizational unit in terms of how it may use a centrally allocated budget for the services it is expected to deliver. Therefore, the need for the central management of such organisations arises to develop a system which incentivizes the local management of each unit to act in a way which optimizes performance (e.g. profits) for the organization as a whole. In particular, an individual organizational unit may choose to maximize its own efficiency which may not be optimal for the organization as a whole in terms of resource use relative to outcomes. For example, a redistribution of resources among operating units may lead to better performance for the organisation as a whole.

One of the key methods deployed for measuring the relative efficiencies of a set of homogeneous operating units or "Decision Making Units" (DMUs) is Data Envelopment Analysis (DEA). This method has generally been developed for the case where the DMUs are independent rather than under some form of central management (e.g., Boussofiane et al. 1991, Thanassoulis 2001; Bogetoft 2013). However, DEA models have also been proposed for the case where the operating units are functioning under a centralised management (e.g. Athanassopoulos 1995; Lozano and Villa 2004; Varmaz et al. 2013). This paper adds to this literature by introducing a method for incentivizing DMUs to act optimally for the overarching central organization. The method takes the approach developed by Varmaz et al. (2013) as its starting point. We highlight some drawbacks of their approach and propose a modified system which not only overcomes those drawbacks but also enhances the functionality of the DEA-based incentives system.

DEA based approaches to managing the performance of multi-unit organisations have been developed so as to address the degree of autonomy enjoyed by the operating units. In this respect, we can discern two broad categories of units: those with limited and those with an enhanced degree of autonomy. For DMUs with limited autonomy, DEA approaches to measuring and managing performance depend on the assumed ease of transferring inputs or re-allocating output requirements between units. Where unimpeded re-allocation of inputs or transfer of outputs is possible, approaches such as those proposed by Athanassopoulos (1995) as well as Lozano and Villa (2004) can be used.

On the other hand, when there exist inflexibilities which make it difficult to reallocate inputs or transfer outputs among the DMUs, approaches such as those proposed by Nesterenko and Zelenyuk (2007) and Asmild et al. (2009) can be used. (See also Fang 2013 and Mar-Molinero et al. 2014 on the re-allocation of inputs and transfer of outputs in centrally managed multi-unit organisations.)

DMUs with enhanced autonomy are those addressed by this paper. Such DMUs have control over their resources, products or services. Central management (the central decision maker or regulator) is responsible only for determining the strategic direction of the organisation as a whole, making policy decisions and monitoring but not micro-managing the activities of the DMUs (Afsharian and Ahn 2014). The central management may want to "induce continuous efficiency improvements, organizational learning and transfer of knowledge with a minimum of control exercised" (Agrell et al. 2002). One approach to measuring and managing performance of such loosely centralised (in effect decentralised organisations) can be found in Varmaz et al. (2013) which is the starting point for this paper.

Varmaz et al. (2013) were the first to propose a DEA-based incentive system for managing the performance of loosely centralised multi-unit organisations. In their 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; alternatively, it is desired to maximize the overall output production by all DMUs given the aggregated inputs they use. To operationalize their system, they modify the centralized resource allocation DEA model (CRA-DEA) proposed by Lozano and Villa (2004). The modification draws among others from the DEA-based incentive mechanism in the context of regulation proposed by Bogetoft (1997). The original framework of this regulatory method – which has its roots in the seminal work of Shleifer (1985) – applies conventional DEA models, i.e. those where the aim is to improve the performance of each unit independently. Varmaz et al. (2013) adapted the framework of the CRA-DEA model of Lozano and Villa (2004) so that it yields a "super-efficiency" measure for each unit. This super-efficiency measure features within the incentive mechanism proposed by Varmaz et al (2013).

The concept of super-efficiency was first proposed by Andersen and Petersen (1993) in order to improve the discrimination between efficient units in DEA. In their approach, the DMU under evaluation is assessed relative to the rest of the DMUs itself being excluded from the comparator units. As a consequence, while inefficient DMUs receive the same efficiency score whether or not they themselves can feature in the comparator set of units, efficient DMUs may have efficiency scores greater than one relative to the most efficient of the rest of the units. The resulting level of

"super-efficiency" leads to discrimination between efficient units. To this effect, super-efficiency measures have been applied for a variety of purposes in the framework of efficiency measurement, such as for incentive regulation (see, e.g., Bogetoft 1997), acceptance decision rules (see, e.g., Seiford and Zhu 1998), detecting exceptional pupils (see, e.g., Thanassoulis 1999), sensitivity analysis in DEA (see, e.g., Zhu 2001), ranking efficient units (see, e.g., Chen 2004) and outlier identification (see, e.g., Banker and Chang 2006).

Varmaz et al. (2013) make use of the concept of super-efficiency as in Bogetoft et al. (1997) in order to devise an incentive compensation formula for the case of centrally managed multi-unit organisations. They modify the CRA-DEA model proposed by Lozano and Villa (2004) in order to arrive at a variant of the super-efficiency measure. The resulting super-efficiency measure is then used within an incentives formula so that inefficient units are encouraged to become more efficient and those with super-efficiency above 100% are incentivised by "rewards" compatible with their level of super-efficiency. As it will be shown in this paper, their approach can lead to counter-intuitive results, incompatible with incentivizing units to improve their performance. The key pitfall in their approach is that their so-called super-efficiency measure does not capture appropriately the impact of a unit on the system as a whole because the system is not defined in a stable manner.

In order to remedy this deficiency, we adapt and extend their approach. We redefine the super-efficiency measure and integrate it in a formula which reflects more accurately how the efficient frontier of the system of units varies with and without each unit under consideration. This in turn leads to a more appropriate level of incentives for each unit. In addition, we highlight that it is possible for units to be "joint super-efficient" if they are located close to each other (in terms of input-output levels). Approaches based on super-efficiency measured by excluding from the referent set one DMU at a time can miss the super-efficiency of jointly super-efficient units. We suggest a practical approach for overcoming this problem so that units which are jointly super-efficient can have their individual super-efficiency identified. This in turn can again lead to more accurate levels of incentives where super-efficiency features in the incentive formula.

The use of super-efficiency in the context of incentive regulation has been researched extensively (see, e.g., Bogetoft 1997, Agrell et al. 2005). A good theoretical foundation of the approach can be found in Bogetoft (1994). In the context of the centralized multi-unit organisations which are the subject of this paper, the problem is modelled in a principal-agent context. The principal (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 it is charged with delivering. This leads to an asymmetry

of information which can be exploited by the agent to extract rents, i.e. the agent finds it takes effort to be cost-efficient and so tends to slacken effort (extract rent). The principal on the other hand wishes to reduce this rent by incentivizing the agents to reveal information that leads to cost-efficient behavior by the agents. One means at the disposal of the principal is the revealed information of the operating costs of all agents. By using methods such as DEA it is possible to arrive at target efficient costs for each agent. The agent is then incentivized (compelled if inefficient) to move towards (not necessarily to) the efficient cost level. However, this approach breaks down as an incentive mechanism for units that are already efficient. They will have no incentive to reveal further efficient cost levels which will lead to a ratchet effect as new efficient cost norms will be incorporated in the compensation formula. It is here that the notion of super-efficiency comes in (see Agrell et al. 2005). The exclusion of a unit from the units used to derive the minimum cost norm for its own compensation eliminates the ratchet effect. Our paper makes use of the concept of super-efficiency in proposing a compensation scheme in the context of a multi-unit organization where the principal is the central management and the agents are the units comprising the organization.

The paper proceeds as follows: Section 2 presents a brief overview of the use of DEA for incentivizing operating units. In Section 3, the approach of Varmaz et al. (2013) is revisited and its drawbacks are identified. In Section 4, we present a new approach for incentivizing the units of a centrally managed multi-unit organisation with respect to their performance. Section 5 extends the approach of Section 4 to cater for cases where the level of super-efficiency of certain units may not be identified due to "joint super-efficiency" of units. Section 6 illustrates our approach using data from a small German retail bank. Section 7 concludes the paper.

2. Overview of the Use of DEA for Incentivizing Operating Units

Let us assume that we have a centrally managed multi-unit organization in which the central management (henceforth "regulator") oversees n agents (i.e. DMUs) who may benefit from a natural monopoly or pre-given rights and flexibilities in producing certain products and/or services. Let $X_j = (x_{1j}, x_{2j}, ..., x_{mj}) \in \Re_+^m$ and $Y_j = (y_{1j}, y_{2j}, ..., y_{sj}) \in \Re_+^s$ be non-zero vectors which quantify the level of inputs and outputs of DMU$_j$ (j=1,…,n). The regulator seeks to incentivise the agents to better performance by controlling their budget b_j (j=1,…,n). However, it is assumed that there is asymmetry of information between the local, autonomous units and the regulator as to the "technology" by which the resources or inputs used by the units are converted to outputs. The regulator can use revealed ex post data on the inputs (X) and outputs (Y) of the units to determine

the budget b_j. In DEA, the technology operated by the units is represented by a production possibility set (PPS) as follows:

$$PPS = \left\{ (X,Y) \in \Re_+^m \times \Re_+^s \mid X \text{ can produce } Y \right\}. \tag{1}$$

Drawing of the work of Shleifer (1985), Bogetoft (1997) proposed the following DEA-based incentive formula by which an optimal compensation plan for DMUp – indicated by b_p – is obtained:

$$b_p = c_p + \rho_p (\theta_p - 1) c_p, \qquad p = 1,...,n. \tag{2}$$

In this formula, c_p is the observed historical cost of the operations of DMU$_p$. θ_p quantifies the efficiency of DMU$_p$ which is obtained by an appropriate DEA model, using the input-output data (X,Y) observed at the operating units. Thus, θ_p represents the fraction of cost c_p that the activities of DMU$_p$ would actually justify if it had been operating as efficiently as benchmark units. Therefore, $(1-\theta_p)$ is the fraction of c_p available for saving. The parameter ρ_p is subjective and moderates the savings fraction ($\theta_p - 1$) imposed on DMU$_p$. Note that we have $(\theta_p - 1) < 1$ when a unit is inefficient while $\rho_p < 1$ would allow for uncertainties in the data such as special factors at play concerning the unit or other considerations which would hamper the unit from saving in full the amounts reflected in $(1-\theta_p)$. As Agrell et al. (2005) note, ρ represents "the power of the incentive scheme". It determines how much of a short fall from efficient cost a unit is forced to save or, as we will see below, when $\theta_p > 1$, ρ reflects what bonus the unit gets to keep out of the amount by which its costs are below benchmark costs.

In the original framework of the above DEA-based reimbursement scheme, conventional DEA models have mainly been used (see, e.g., Bogetoft 2013), such as the BCC DEA model proposed by Banker et al. (1984):

$$\theta_p = \max \left\{ \sum_{r=1}^{s} u_{rp} y_{rp} + \mu \;\middle|\; \sum_{i=1}^{m} v_{ip} x_{ip} = 1, \right.$$

$$\sum_{r=1}^{s} u_{rp} y_{rj} - \sum_{i=1}^{m} v_{ip} x_{ij} + \mu \leq 0, \qquad\qquad j = 1,...,n$$

$$\left. v_{ip} \geq 0, \;\; i=1,...,m; \;\; u_{rp} \geq 0, \;\; r=1,...,s; \;\; \mu \text{ free in sign} \right\}. \tag{3}$$

In model (3), (u_{rp}, v_{ip}) stand for the weights (multipliers) of outputs and inputs respectively, θ_p represents the relative efficiency score of DMU$_p$ and μ reflects the returns to scale (RTS) (increasing, constant or decreasing) of the unit p. Depending on the context, one could impose, e.g., constant, non-decreasing or non-increasing returns to scale by setting μ as zero, non-negative or non-positive, respectively. Furthermore, (3) reflects an input-oriented DEA model in the sense that targets are sought which minimize inputs (e.g. operating costs), controlling for output levels. The model can be readily converted to an output-oriented one (see, e.g., Thanassoulis 2001).

The above model determines individual efficiency scores which have a range between zero and one. On this basis and according to (2), the costs of an agent with a relative efficiency score of one will be reimbursed fully while the inefficient agents are not fully reimbursed. However, in order to give incentives to the efficient agents, Bogetoft (1997) proposed modifying model (3) by means of the super-efficiency concept of Andersen and Petersen (1993). The basic idea is to compare DMU$_p$ under evaluation with all other units in the sample except itself so that DMU$_p$ cannot influence its own benchmark. This modification can be incorporated in model (3) (see Andersen and Petersen 1993) by replacing the constraints $j=1,\ldots,n$ by those in (4).

$$\sum_{r=1}^{s} u_{rp} y_{rj} - \sum_{i=1}^{m} v_{ip} x_{ij} + \mu \leq 0, \qquad j = 1,\ldots,n; \;\; j \neq p. \qquad (4)$$

The resulting model will provide the same scores as before for the inefficient units, while the efficient ones will normally have efficiency scores greater than one. Hence, utilizing these super-efficiency scores in (2), agents with poor performance are not fully reimbursed, while performers with super efficiency $\theta_p^s > 1$ can be awarded budgets above c_p (depending on the ρ_p value chosen) to incentivize them to better performance in future.

3. The Incentivisation System Proposed by Varmaz (2013)

Unlike in the framework addressed by Bogetoft (1997), Varmaz et al. (2013) address specifically the objective of improving the efficiency of the whole system of units rather than improve efficiency at unit level. Instead of conventional DEA models such as the one in (3), they use the CRA-DEA model proposed by Lozano and Villa (2004) as a vehicle for determining resource allocation to units. This model, whose mathematical structure has recently been simplified by Mar-Molinero et al. (2014), is as follows:

$$Eff_o = \max\left\{\sum_{j=1}^{n}\sum_{r=1}^{s}u_r y_{rj} + n\mu \;\middle|\; \sum_{j=1}^{n}\sum_{i=1}^{m}v_i x_{ij} = 1,\right.$$

$$\sum_{r=1}^{s}u_r y_{rj} - \sum_{i=1}^{m}v_i x_{ij} + \mu \leq 0, \qquad\qquad j = 1,...,n$$

$$\left.v_i \geq 0, \;\; i=1,...,m; \;\; u_r \geq 0, \;\; r=1,...,s; \;\; \mu \; free \; in \; sign\right\}. \tag{5}$$

The model in (5) identifies output weights u_r, $r=1,...s$ and input weights v_i, $i=1,...,m$ which are applied to the input-output levels of all units $j=1,...,n$. In the objective function and the normalization constraint, the aggregate levels of each output and of each input across the units are used. The constraints $j=1,...,n$ are those of a classical input-oriented DEA model under VRS, ensuring that no individual unit, when its inputs and outputs are evaluated by the above common set of weights, has an efficiency above the nominal 1. It is possible for some of the units $j=1,...,n$ to be given by the common input and output weights a negative efficiency value, should the value of μ be highly negative (see Soares de Mello et al. 2013). However, it can be readily seen that a negative value for the overall efficiency Eff_o though mathematically feasible, it will never be optimal.

The model in (5) provides a single overall efficiency score Eff_o for the entire group of units or equivalently for a virtual unit that has the mean value of each input and output computed across all units. However, the reimbursement formula in (2) requires individual efficiency scores to determine the respective compensation level for each unit. Therefore, Varmaz et al. (2013) proposed a measure of efficiency which determines the individual "contribution" of each unit to the efficiency of the whole system. This measure is based on a comparison of the efficiency of the entire system – obtained from the above model in (5) (denoted by Eff_o) – with the efficiency score of the system where the unit under evaluation is excluded entirely from the set of observations in the analysis (indicated by $Eff_o^{\neq p}$). Thus, $Eff_o^{\neq p}$ can be computed by means of the model in (6):

$$Eff_o^{\neq p} = \max\left\{\sum_{\substack{j=1\\j\neq p}}^{n}\sum_{r=1}^{s}u_r y_{rj} + n\mu \;\middle|\; \sum_{\substack{j=1\\j\neq p}}^{n}\sum_{i=1}^{m}v_i x_{ij} = 1,\right.$$

$$\sum_{r=1}^{s}u_r y_{rj} - \sum_{i=1}^{m}v_i x_{ij} + \mu \leq 0, \qquad j = 1,...,n, \; j \neq p$$

$$\left.v_i \geq 0, \;\; i=1,...,m; \;\; u_r \geq 0, \;\; r=1,...,s; \;\; \mu \; free \; in \; sign\right\}. \tag{6}$$

Once the value of $Eff_o^{\neq p}$ is known, the ratio in (7) is computed as[2]

$$\tilde{\theta}_p = \frac{Eff_o}{Eff_o^{\neq p}}. \tag{7}$$

$\tilde{\theta}_p$ reflects the difference in the overall efficiency score between including and excluding the unit under evaluation p from the analysis. Varmaz et al. (2013) argue that the higher value of $\tilde{\theta}_p$, the higher the individual "contribution" of DMU_p to the overall efficiency of the system of operating units. It should be noted that – due to the nature of the efficiency measure in (6) – $Eff_o^{\neq p}$ and $\tilde{\theta}_p$ can also take values above one.

Although the core idea behind the approach proposed by Varmaz et al. (2013) is interesting, it is questionable whether the "contribution" measure in (7) can determine a suitable incentive level for each unit. In order to illustrate the problem, consider the simple case depicted in Figure 1(a), where there exist eight units with two inputs and one output, i.e. all units are assumed to have the same level of output.

Figure 1. Representation of the production possibility set within the approach of Varmaz et al. (2013)
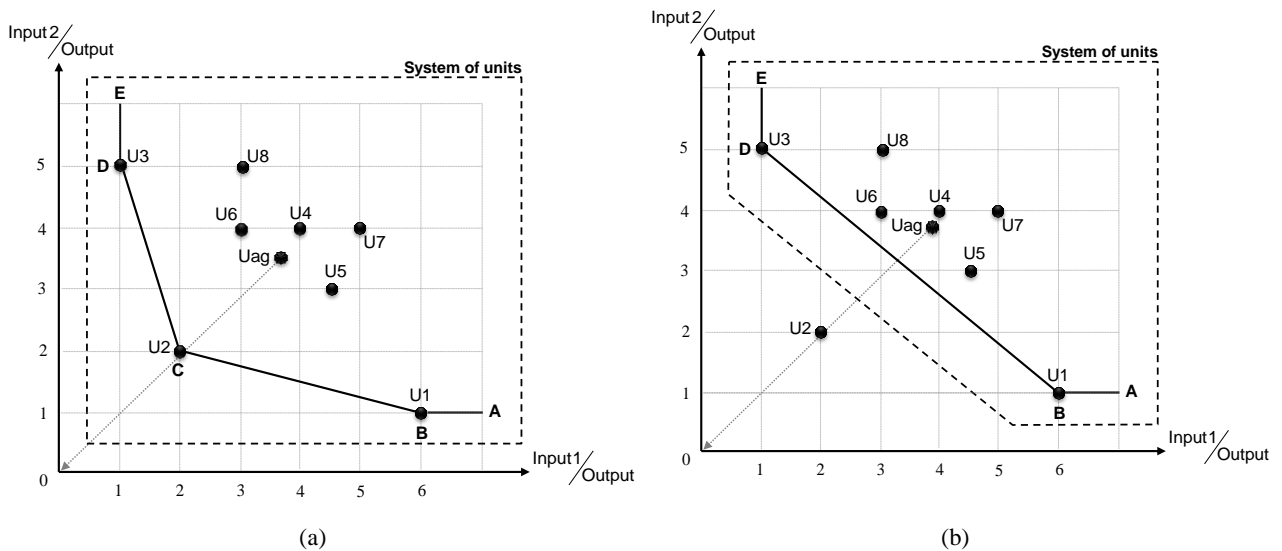
(a)   (b)

Applying the model in (5), the overall efficiency of the whole system, depicted as U_{ag} in Figure 1, will be 0.569, i.e. $Eff_o = 0.569$. The corresponding PPS in Figure 1(a) is bounded by ABCDE. In order to determine the individual contribution of each unit to the overall efficiency of the whole system according to Varmaz et al. (2013), model (6) has to be solved for each unit. The results of this model and the corresponding values of measures used in the Varmaz et. al. (2013) approach are given in Table 1.

Table 1. Results for the example determined by the approach of Varmaz et al. (2013)

Unit	θ_p	$Eff_o^{\neq p}$	$\tilde{\theta}_p$
1	1.000	0.593	0.961
2	1.000	0.860	0.662
3	1.000	0.586	0.972
4	0.500	0.581	0.980
5	0.606	0.577	0.986
6	0.615	0.576	0.988
7	0.476	0.593	0.961
8	0.571	0.596	0.956

As an example, let us focus on the result for unit #2, i.e. U_2 in Figure 1(a). The efficiency score $Eff_o^{\neq p}$ of this unit obtained from model (6) is 0.860. This shows that the efficiency score of the whole system increases significantly when this unit is excluded entirely from the analysis, c.f. $Eff_o = 0.569$ and $Eff_o^{\neq 2} = 0.860$. The measure $\tilde{\theta}_2$ (see (7) of this unit) with a value less than one (i.e. 0.662) also captures its negative contribution to the overall efficiency of the whole system. However, as can be seen in Figure 1(a), this unit is not only efficient considering any conventional DEA model – like model (3) which reports an efficiency score of $\theta_2 = 1$ (see the second column of Table 1) –, but also uses the resources in a balanced way compared to the other units (see, e.g., Ahn et al. 2012). Moreover, unit #2 is a benchmark unit when measuring the efficiency of the whole system by model (5), attesting to the fact that other units should emulate unit #2 in order for the performance of the whole system to improve. Consequently, it would be very difficult to accept that this unit should not be compensated in the framework of an incentives formula as proposed by Varmaz et al. (2013). Such a counter-intuitive result is not limited to this example.

Varmaz at al. (2013, p. 113) themselves introduced a requirement for an appropriate performance measure, stating 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". A look at the "super-efficiency" scores in the last column of Table 1 reveals that there is no unit with a positive contribution to the

overall efficiency of the whole system, i.e. all values are less than one. Due to this result, all units are considered inefficient so that their costs are not fully reimbursed. This is obviously inconsistent with the definition of relative efficiency itself as there is no unit with a full efficiency score in the results. These results also contradict the concept of the "super-efficiency" measure and the above statement by which high-powered incentives should be given to efficient units. It must therefore be concluded that the approach of Varmaz at al. (2013) cannot always incentivize units to improved performance under centralized management.

The reason for the counter-intuitive results above can be explained by looking at the ratio in (7). Eff_o and $Eff_o^{\neq p}$ are both "relative" efficiency scores ranging from zero to one. As can be seen in Figure 1(a), for determining Eff_o as the overall efficiency score, all units are included in the relative efficiency measurement system. The aggregate unit expressed in mean values of inputs and output is depicted in U_{ag}. However, $Eff_o^{\neq p}$ is computed by means of all units except the unit under evaluation. As it is shown in Figure 1(b), for determining $Eff_o^{\neq 2}$, e.g., unit #2 is excluded entirely both from the potential referent units and from the computation of the aggregate levels of the inputs and outputs of the system of units. The resulting PPS, depicted in Figure 1(b), is now bounded by ABDE while the aggregate unit has input-output levels at U_{ag}. It is clear now that $Eff_o^{\neq p}$ and Eff_o cannot capture appropriately the impact of unit U_2 on the system of units because their ratio conflates the change both in the efficient boundary and in the definition of the system of units when unit U_2 is removed. This can be seen when comparing the positions of the efficient boundary and of unit U_{ag} in Figures 1(a) and 1(b) where two different units are reflected on two different efficient boundaries. Consequently, a comparison of the two resulting scores cannot be readily interpreted.

In the next section, we propose a modification of the approach of Varmaz et al. (2013) so that the foregoing efficiency measures can be compared more readily in the framework of an incentivisation formula under centralized management.

4. The Proposed Incentivisation Method for Centrally Managed Organisations

Let us revisit the original reimbursement scheme given in (2) in conjunction with model (3). The model has to be run n times in order to determine the maximum efficiency score for each unit and the corresponding optimal input-output weights (u_{rj}, v_{ij}) for each unit j. This process implies that each unit may use different implicit values for inputs and outputs for calculating its efficiency. This flexibility reflects an important principle of DEA models when DMUs operate independently, each

one according to its own priorities. However, it is not a desired approach in centralized management systems, such as those addressed in this paper, where the fundamental objective is improving the performance of the whole system rather than that of individual units.

The reason is that selecting the weights in such a manner may allow the DMUs to focus on different goals and strategies which might be inconsistent with the central decision maker's preferences, e.g., organizational goals and strategies (for more details, see, e.g., Chen and Zhu 2011). In the worst case, this flexibility of conventional DEA models may even allow a DMU in the system to assign very low implicit values to some inputs/outputs in order to exclude them from the analysis. This is inappropriate, especially under a centralized management which aims to improve the system's performance. Instead, the central management may select the input-output factors and measure the efficiency periodically in order to direct the units towards the organizational goals and strategies. Thereby, it wants to preserve the consistency across the units and keep the overall focus on the strategic direction of the organization. In such situations, the central management can apply a common set of preferences to improve the performance of the whole system.

Several authors have shown that such a structure can be modelled by deriving a common basis – in terms of a common set of input and output multipliers – for measuring efficiency in centrally managed organisations. Examples include common-weights DEA models proposed by Roll et al. (1991), Roll and Golany (1993), Cook et al. (2004), Kao and Hung (2005), Cook and Zhu (2007) as well as Zohrehbandian et al. (2010). A close look at the structure of model (5) reveals that it also belongs to the above outlined family of the common-weights DEA models (see Lozano and Villa 2004, p. 149). As in the case of the other members of this family, the model is therefore capable of determining a common set of input-output multipliers as a by-product of measuring overall efficiency. The respective set of weights can be used for measuring the efficiency of all units on a conjoint basis. More precisely, after having computed a common set of input-output multipliers by model (5) – indicated by (u_r^*, v_i^*) –, one can derive an individual common-weights efficiency score of DMU_p $(p=1,…,n)$ by the following ratio of efficiency (RE):

$$RE_p^{CM} = \frac{\sum_{r=1}^{s} u_r^* y_{rp} + \mu^*}{\sum_{i=1}^{m} v_i^* x_{ip}} . \tag{8}$$

The results are marked by a superscript *"CM"* to emphasize that the individual efficiency scores are now computed within the proposed framework under centralized management and based on the common-weights DEA model (5). We note here, however, that the optimal solution of model (5) (though rarely) may not be unique. Thus, a unique set of individual common-weights efficiency scores may not be obtained either. The model in (9) is proposed for arriving at the common-weights efficiency θ_p^{CM} for DMU p:

$$\theta_p^{CM} = \max \left\{ \sum_{r=1}^{s} u_{rp} y_{rp} + \mu \; \middle| \; \begin{array}{l} \sum_{i=1}^{m} v_{ip} x_{ip} = 1 \\[2mm] \sum_{r=1}^{s} u_{rp} y_{rj} - \sum_{i=1}^{m} v_{ip} x_{ij} + \mu \leq 0, \qquad j = 1,...,n \\[2mm] \sum_{j=1}^{n} \sum_{r=1}^{s} u_{rp} y_{rj} + n\mu - Eff_o \sum_{j=1}^{n} \sum_{i=1}^{m} v_{ip} x_{ij} = 0 \\[2mm] v_{ip} \geq 0, \; i = 1,...,m; \; u_{rp} \geq 0, \; r = 1,...,s; \; \mu \; free \; in \; sign \end{array} \right\}. \qquad (9)$$

This model in (9) seeks the set of multipliers which maximize the efficiency score of DMU$_p$ (p=1,…,n), while Eff_o – the overall efficiency score of the entire system computed by model (5) – is fixed at its previous optimal value and has been added as a constraint. Hence, θ_p^{CM} represents the efficiency score of DMU$_p$ in the best possible light under centralized management.

As in the case of model (5), – and most other common-weights DEA models – the efficiency θ_p^{CM} yielded by (9) under VRS may in certain cases take negative values. This would be the case when μ is negative and has an absolute value large enough to render the optimal objective function value in (9) negative. However, this will happen less often than in the case when the common-weights are directly extracted from model (5) because model (9) will maximise the μ value compatible with the value of Eff_o. Where a negative value of θ_p^{CM} does result, one of the procedures outlined in Soares de Mello et al. (2013) can be used to resolve the problem. For simplicity, we shall in the remainder of this paper ignore the case of a negative value for θ_p^{CM}.

Theorem 1. *The efficiency score of DMU$_p$ determined by the proposed centralized model (9), represented by θ_p^{CM}, never exceeds θ_p, the efficiency score of this unit under decentralized management computed by model (3).*

Proof. Let us assume that (u_{rp}, v_{ip}) is a feasible solution to the model in (9). Therefore, (u_{rp}, v_{ip}) satisfies each of its constraints. This implies that (u_{rp}, v_{ip}) is a feasible solution to model (3) as it also satisfies the constraints in this model. Thus $\theta_p^{CM} \leq \theta_p$. ∎

Although the model in (9) can determine a set of relative efficiency scores θ_p^{CM} consistent with optimizing the performance of the system as a whole, it can only provide efficiency values within a range of zero to one. In line with Bogetoft (1997) and Varmaz et al. (2013), our proposed framework needs to also be capable of yielding efficiencies above 1 for units that are good performers and we want to incentivize them to perform better still. In order to meet this requirement, we propose that the model in (10) be solved for each unit p under evaluation:

$$Eff_o^p = \max \left\{ \sum_{j=1}^{n} \sum_{r=1}^{s} u_r y_{rj} + n\mu \; \middle| \; \sum_{j=1}^{n} \sum_{i=1}^{m} v_i x_{ij} = 1, \right.$$
$$\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} + \mu \leq 0, \qquad j = 1,...,n, \quad j \neq p \right\}$$
$$\left. v_i \geq \varepsilon, \; i = 1,...,m; \; u_r \geq \varepsilon, \; r = 1,...,s; \; \mu \text{ free in sign} \right\}.$$

(10)

This model is a modification of model (5) in such a way that the DMU under evaluation itself is excluded from candidate boundary units. This is reflected in the constraints $j=1...n, j \neq p$ in (10). Note however that unit p is not excluded from the aggregation of units of the system, as reflected in the objective function. This is a crucial difference from the model proposed by Varmaz et al. (2013) in that we retain constant the definition of the aggregate unit representing the system across all instances of model (10). This provides a constant reference point for capturing the changes in the efficient boundary as each unit p in turn is taken out of the constraints to model (10). Another difference from the model proposed by Varmaz et al. (2013) is that in model (10) we restrict the input/output weights (u_r, v_i) to be strictly positive rather than non-negative. The requirement for strictly positive weights is standard in DEA models as it ensures that models such as the one in (10) lead to Pareto efficient solutions (e.g. see also Thanassoulis 2001, model (3.4), for more details).

It should be noted that the model in (10) – like other standard VRS DEA models which are based on the super-efficiency concept – might be infeasible. This can occur when the aggregate DMU under evaluation is not enveloped by the boundary formed by all bar the excluded DMU_p. This occurrence though theoretically possible should be rare since in essence the aggregate unit is seen by the assessment model as having the average level of each input and each output observed and so it

should normally be enveloped. In cases where model (10) yields no feasible solution, one approach that can be adopted to arrive at a super-efficiency for unit p is to change the orientation of the model (e.g. from input to output orientation). This will in most cases lead to a feasible solution which still can reflect the impact of unit p on the system. In the highly unusual case where the model is infeasible in both input and the output orientation we in effect have insufficient information to offer such unusual units incentive through model (10) (see also Agrell et al. 2005 on this point). Such units would need to be considered by the central management as special cases.

Remark 1. *The definition of model (10) inherits the original concept of the super-efficiency measure outlined in Section 2.1 so that $Eff_o \leq Eff_o^{\,p}$, as model (10) has one less constraint than the model in (5).*

We can now use the measure in (11) in the framework of the incentivisation formula in (2).

$$\tilde{\theta}_p^{CM} = \frac{Eff_o^{\,p}}{Eff_o} \times \theta_p^{CM} \tag{11}$$

Eff_o, θ_p^{CM} and $Eff_o^{\,p}$ are computed by means of (5), (9) and (10), respectively.

The following scenarios can occur in respect of the value in (11):

1. The unit under evaluation p is inefficient, i.e. θ_p^{CM} determined by (9) represents a value less than one. In this case, $Eff_o^{\,p}$ equals Eff_o automatically and so $\tilde{\theta}_p^{CM}$ equals θ_p^{CM}. As this value is less than one, such units are not reimbursed fully but are required to make certain savings to improve efficiency in line with the formula in (2).

2. The unit p under evaluation is a benchmark unit and so θ_p^{CM} is one. Then, two situations may arise:

2.1. The exclusion of unit p does not affect the assessed performance of the whole system. In this case, $Eff_o^{\,p}$ also equals Eff_o, which results in $\tilde{\theta}_p^{CM} = \theta_p^{CM} = 1$. This will be the case when in rare circumstances 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). Units of this type may not be required to make savings but will also not be offered budgets in excess of expenditure unless they are identified as being "jointly super-efficient" in the manner outlined

in Section 6. In that case, the procedure in Section 6 can lead to compensation beyond their level of expenditure.

2.2. The exclusion of the unit under evaluation affects the assessed performance of the whole system. In this case, Eff_o^p will be greater than Eff_o , which leads to $\tilde{\theta}_p^{CM} > 1$. We propose that this $\tilde{\theta}_p^{CM}$ be used within the incentivisation formula (2).

Remark 2. *As the model in (10) retains the same efficiency scores as before for the inefficient units (see scenario 1 in which $Eff_o^p = Eff_o$ and $\tilde{\theta}_p^{CM} = \theta_p^{CM}$), the model in (10) and consequently the formula (11) need to be calculated only for units whose $\theta_p^{CM} = 1$ (see scenario 2 above).*

Thus in the approach being proposed inefficient units where $\tilde{\theta}_p^{CM} < 1$ will not be fully compensated and hence they will be encouraged to make efficiency savings. In contrast, for efficient units characterized by $\tilde{\theta}_p^{CM} > 1$, the incentivisation formula in (2) gives higher budgets for a period of time than their projected operating costs c_p. This would prove an incentive to higher efficiency because the higher the super-efficiency a unit registers the larger the budget that can be awarded to be used going forward at the discretion of local management. In practice, the central management has the opportunity to control the incentive level through the parameter ρ in (2), and potentially the period of time over which the incentive is spread. One consideration in this context would be to ensure that the incentive levels will maintain the units concerned as benchmarks relative to other units.

It is important to note that using $\tilde{\theta}_p^{CM}$ within the incentivisation formula in (2) cannot lead to over-compensating a unit that had been inefficient under standard DEA. This is a consequence of Theorem 2.

Theorem 2. *It is the case that $\tilde{\theta}_p^{CM} \leq \theta_p^s$, where θ_p^s is the super-efficiency of unit p as derived using the Andersen and Petersen 1993 approach.*

Proof. See the Appendix.

Theorem 2 states that the derived efficiency $\tilde{\theta}_p^{CM}$ for unit p to be used in the incentivisation formula in (2) cannot exceed its super-efficiency under decentralized management θ_p^s (as derived using the Andersen and Petersen 1993 approach). Thus, the incentivisation formula in (2), when $\tilde{\theta}_p^{CM}$ is used

within it, will never over-compensate a unit that is inefficient in a decentralized context. As we might expect, a unit p that is not efficient even when it pursues its own best interests cannot be offered an incentive to become even more inefficient when pursuing global system objectives under centralized management. Furthermore, even when a unit p is efficient in our approach, it will still not be over-compensated if its removal from the efficient boundary does not affect the location of that boundary. In this case, we have $\tilde{\theta}_p^{CM} = \theta_p^{CM} = 1$, as can be seen in the proof in the Appendix.

Thus, our approach only offers an over-compensation to a unit when its location in terms of input-output levels is material for the location of the PPS frontier. This property of our approach is in contrast with that of Varmaz et al. (2013), where inefficient units in a decentralized context can be over-compensated while the reverse is also possible, where super-efficient units in a decentralized context are not incentivized through over-compensation (see Table 1).

We conclude this section by a graphical illustration of the incentivisation approach we propose. Consider again the example given in Section 3. Applying model (5), the efficiency of the whole system based on model (3) is 0.569, denoted by $Eff_o = 0.569$. The corresponding PPS in Figure 2(a) is bounded by ABCDE.

Figure 2. Representation of the production possibility set within the proposed approach

(a)   (b)
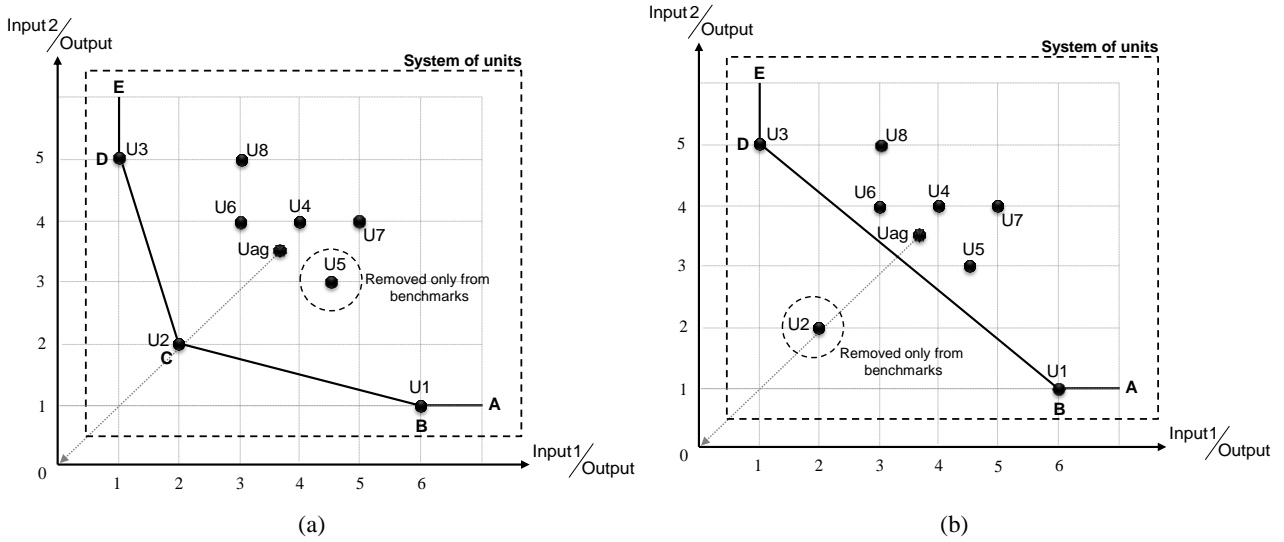
Solving the model in (9) provides the individual efficiency scores θ_p^{CM} under centralized management, which are reported in the second column of Table 2. Solving next model (10) for each unit, we obtain the efficiencies Eff_o^p, and using them in (11), we obtain the adjusted θ_p^{CM} denoted by $\tilde{\theta}_p^{CM}$. The results are given in the last two columns of Table 2.

Table 2. Results for the example determined by our proposed approach

Unit	θ_p^{CM}	$Eff_o^{\,p}$	$\tilde{\theta}_p^{CM}$
1	1.000	0.571	1.004
2	1.000	0.913	1.604
3	0.476	0.569	0.476
4	0.500	0.569	0.500
5	0.606	0.569	0.606
6	0.526	0.569	0.526
7	0.476	0.569	0.476
8	0.435	0.569	0.435

Let us look, e.g., at unit #5 depicted as U_5 in Figure 2(a). The efficiency score of this unit obtained via the model in (9) is $\theta_5^{CM} = 0.606$. As the unit is inefficient, its exclusion as candidate benchmark unit in the model in (10) does not change the value of efficiency estimated for the whole system (see Figure 2(a)). This means that Eff_o^5 equals Eff_o ($= 0.569$), and so $\tilde{\theta}_5^{CM} = 0.606$. This efficiency used in (2) would imply the unit needs to make savings.

In contrast, if we take unit #2 and assess it using model (9), the resulting efficiency is $\theta_p^{CM} = 1$. Assessing the performance of the whole system by model (10), the result is $Eff_o^2 = 0.913$. This value implies that the efficiency score of the whole system increases significantly when unit #2 is excluded from units which serve as benchmarks in (10), c.f. $Eff_o^2 = 0.913$ compared to $Eff_o = 0.569$ when unit 2 is among the benchmarks. Thus $\tilde{\theta}_2^{CM} > 1$, meaning that the unit should be incentivised by the award of a budget in excess of what it had spent. The central management could use the parameter ρ_p in (2) to moderate the fraction ($\tilde{\theta}_2^{CM} - 1$) of c_p that the unit is awarded as discretionary budget over and above their estimated operating costs of c_p.

Note that unit #2 was also assessed earlier by the method proposed by Varmaz et al. (2013) yielding counter-intuitive results unlike those obtained with the method proposed in this paper. The difference can be traced to the way super-efficiencies are computed in the two methods. As can be seen in Figure 2(b), unlike in the Varmaz et al. (2013) approach, our suggested framework does not exclude unit #2 from the aggregate system of units when the super-efficiency of the latter is assessed. It is only excluded from the set of units which can serve as benchmarks. Therefore, the resulting value of "contribution" reflects the original concept of the super-efficiency central to both approaches.

5. A Modified Approach to Cater for Joint Super-Efficiencies

A potential problem with super-efficiency based approaches, including that of Varmaz et al. (2013), is that they may not identify properly the impact of certain units if they have very similar performance to each other, even if jointly they are significantly different from other units in the system. In order to illustrate the issue, consider once again the numerical example in the previous section. Let us now assume that there exists another unit in the system whose operation in terms of input-output levels is very similar to that of unit #2. The entire system is depicted in Figure 3(a) in which the new unit is represented by U_9.

Figure 3. Representation of the production possibility set within the proposed extended approach
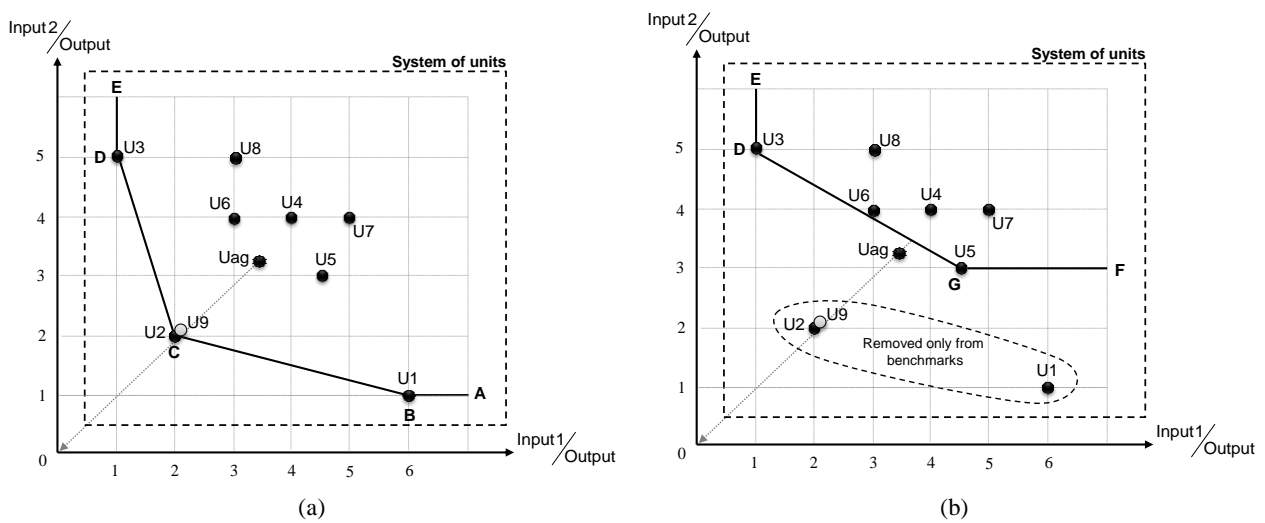
Applying model (5), the overall efficiency of the whole system will be 0.598, i.e. $Eff_o = 0.598$. The corresponding PPS in Figure 3(a) is bounded by ABCDE. Solving the proposed models in (9) and (10), the formula in (11) can then be used to determine the individual efficiency scores and the contribution of each unit to the overall efficiency of the whole system. The results are given in Table 3.

Table 3. Results for the modified example determined by the proposed approach

Unit	θ_p^{CM}	Eff_o^p	$\tilde{\theta}_p^{CM}$
1	1.000	0.600	1.003
2	1.000	0.601	1.005
3	0.476	0.598	0.476
4	0.500	0.598	0.500
5	0.606	0.598	0.606
6	0.526	0.598	0.526
7	0.476	0.598	0.476
8	0.435	0.598	0.435
9	0.995	0.598	0.995

As can be seen in Table 3, the contribution of unit #2 has decreased significantly from the previous value of 1.604 where unit #9 was not included in the example (see Table 2: $\tilde{\theta}_2^{CM} = 1.604$) to the new value of $\tilde{\theta}_2^{CM} = 1.005$. The reason is that excluding unit #2 results in a very similar shape of the PPS by which model (10) determines a super-efficiency value of $Eff_o^2 = 0.601$. In addition, as unit #9 is inefficient (i.e. $\theta_9^{CM} = 0.995$), the exclusion of this unit does not affect the performance of the whole system in model (10), i.e. $Eff_o^9 = Eff_o = 0.598$. Accordingly, the contribution of this unit to the overall performance is represented by a value of $\tilde{\theta}_9^{CM} = 0.995$. This phenomenon has occurred because units #9 and #2 mask the "exceptional" performance of each other. Therefore, two or more neighbouring units in terms of mix of inputs and output levels having similar high performance – even if collectively significantly away from the rest of the units – will not be identified as "significantly" super-efficient, if at all.

This is a technical problem which can be overcome by a form of sensitivity analysis. As a solution, we propose an adaptation of the procedure introduced by Thanassoulis (1999). We first determine initial efficiencies θ_p^{CM}. We then collect within a set $S_{100\text{-}r}$ all the units with an efficiency $\theta_p^{CM} \geq (100 - r)\%$ where 100-r is a user-specified efficiency level which is close enough to 100% to be deemed as 100%. We now assess again the units not in $S_{100\text{-}r}$ adding to the set $S_{100\text{-}r}$ any new units with θ_p^{CM} efficiency at or above (100-r)%. This is done until either there are no more units with θ_p^{CM} efficiency (100-r)% or a significant percentage π (e.g., $\pi = 5\%$) of the full set of units has been placed within the aggregate set $S_{100\text{-}r}$. At that point, we have a final set $FS_{100\text{-}r}$. The units in $FS_{100\text{-}r}$ are potentially super-efficient.

We now re-calculate the adjusted efficiency $\tilde{\theta}_p^{CM}$ values for each one of the units in $FS_{100\text{-}r}$ as follows. For each one of these units, model (10) is recomputed using the complement of the set $FS_{100\text{-}r}$ by replacing in model (10) its constraint with the constraint in (12).

$$\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} + \mu \leq 0, \qquad j = 1,...,n, \quad j \notin FS_{100-r}. \tag{12}$$

Thus, the modified model in (10) measures the efficiency of the whole system while its potential benchmarks – represented by the constraint (12) – do not include the units in $FS_{100\text{-}r}$. In other words, in measuring Eff_o^p for each one of the units in $FS_{100\text{-}r}$, the performance of the whole system is

compared to another system which excludes jointly the units with a θ_p^{CM} efficiency value greater than or equal to (100-*r*)% as we progressively remove super-efficient units. Note that the exclusion is only in respect of being potential frontier units and not from the definition of the aggregate inputs and outputs of the system of units which is stable throughout.

Using – on the data in Table 3 – *r* = 0.5 and π=5% within the foregoing procedure, we deduce that units #1, #2 and #9 belong to S$_{99.5}$. The three units represent over 5% of all units and so FS$_{99.5}$ consists of units #1, #2 and #9. Excluding jointly these units will lead to the PPS labelled as FGDE in Figure 3(b). Hence, this PPS is now used to re-calculate the super-efficiency scores of the units in FS$_{99.5}$. The results are summarized in Table 4.

Table 4. Results for the modified example determined by the procedure to identify jointly super-efficient units

Unit	θ_p^{CM}	$Eff_o^{\,p}$	$\tilde{\theta}_p^{CM}$
1	1.000	1.057	1.768
2	1.000	1.057	1.768
3	0.476	0.598	0.476
4	0.500	0.598	0.500
5	0.606	0.598	0.606
6	0.526	0.598	0.526
7	0.476	0.598	0.476
8	0.435	0.598	0.435
9	0.995	1.057	1.759

As can be seen, the inefficient units out of FS$_{99.5}$ have received the same contribution scores as before, cf. Table 3 and Table 4 for units 3 to 8. However, the contribution values of units in S$_{99.5}$, reported in the last column of Table 4, are all positive and significantly higher than what had been reported in Table 3.

Clearly, the parameters *r* and π are subjective. The *r* conveys the user's sense of what constitutes a close enough distance to 100% efficiency and π conveys the user's view as to what proportion of units achieving a level of efficiency means that efficiency is no longer "exceptional". For example, if 5% or more units exceed a level of attainment, then that level is within reach and an acceptable benchmark rather than exceptional. As *r*% and π% are user-dependent, it makes good sense for the user to carry out a respective sensitivity analysis.

6. An Empirical Illustration Using Data from a German Retail Bank

In order to numerically illustrate the features of our proposed approach and to show its practical relevance, we analyse the performance of 16 branches of a small German retail bank. The data set, which originates from Varmaz et al. (2013), is shown in Table 5.

Table 5. 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 data set comprises two inputs and two outputs which were chosen by Varmaz et al. (2013) drawing on work by Avkiran (2009). The inputs are personnel expenses (PEX) as well as expenses on interest payments (IEX). IEX consist of payments of interest for deposits. In addition to wages, PEX also include payments for employee training and social insurance contributions. Interest income (IIN) and all other income (OIN) are considered as the two outputs in the bank branches. OIN mainly contain the fees earned by branches from securities trading, transactions and insurance sales. (For a more detailed description of these inputs and outputs, see Varmaz et al. 2013).

The models developed in this paper have been encoded in AIMMS, version 3.13. Applied to the data in Table 5, Table 6 summarizes the results from these models. Its second column presents the results obtained by the BCC model (3). Modifying this model according to (4), the corresponding super-efficiency scores θ_p^s are shown in the third column. Seven branches (1, 4, 6, 7, 12, 15 and 16) are reported to be efficient[3]. The single overall efficiency score of the bank with its 16 branches, computed by model (5), is 0.835 (see the fourth column). The next two columns summarize the results obtained by model (6) and formula (7) within the centralized framework proposed by Varmaz et al. (2013). $Eff_o^{\neq p}$ represents the efficiency score of the bank where the pth branch is excluded entirely from the set of observations in the analysis, while $\tilde{\theta}_p$ reports the individual contribution of this branch to the overall efficiency of the whole system of bank branches, derived as the ratio $0.835/Eff_o^{\neq p}$.

---

[3] It should be noted that the first unit was not enveloped in the input orientation (more details about the infeasibility issue of super-efficiency models can be found in, e.g., Seiford and Zhu, 1999; Chen 2005; Chen and Liang 2011; Lee et al. 2011).

Table 6. Results of the various efficiency measures previously defined

Unit	θ_p	θ_p^s	Eff_o	$Eff_o^{\neq p}$	$\tilde{\theta}_p$	θ_p^{CM}	Eff_o^p	$\tilde{\theta}_p^{CM}$
1	1.000	Na.	0.835	0.808	1.033	1.000	0.847	1.014
2	0.981	0.981		0.817	1.022	0.978	0.835	0.978
3	0.818	0.818		0.843	0.991	0.753	0.835	0.753
4	1.000	1.075		0.837	0.998	0.755	0.835	0.755
5	0.699	0.699		0.845	0.988	0.674	0.835	0.674
6	1.000	1.242		0.846	0.987	0.735	0.835	0.735
7	1.000	1.182		0.839	0.995	1.000	0.845	1.012
8	0.837	0.837		0.838	0.996	0.786	0.835	0.786
9	0.892	0.892		0.841	0.993	0.643	0.835	0.643
10	0.831	0.831		0.842	0.992	0.669	0.835	0.669
11	0.923	0.923		0.837	0.997	0.817	0.835	0.817
12	1.000	1.176		0.850	0.982	1.000	0.858	1.027
13	0.807	0.807		0.841	0.993	0.643	0.835	0.643
14	0.864	0.864		0.835	1.000	0.835	0.835	0.835
15	1.000	1.037		0.840	0.993	0.590	0.835	0.590
16	1.000	1.262		0.833	1.002	0.924	0.835	0.924

Taking a closer look at $\tilde{\theta}_p$, four branches (1, 2, 14 and 16) are seen to have a contribution of 1 or higher to the overall efficiency, and so deserving a budget beyond the expenditure they report. In particular, branches #2 and #14 have contributions of 1.022 and 1.000, respectively. However, the corresponding conventional efficiency score of these branches, reflected in θ_p (or θ_p^s), are 0.981 and 0.864, respectively. In both cases, a branch which is inefficient under the decentralized framework is being "rewarded" in a centralized scenario. This result is counter intuitive, as noted earlier.

Under the decentralized management scenario, the fundamental objective of the benchmarking system is defined as improving the performance of each individual unit, independently. This objective is reflected in the mathematical structure of the decentralized framework whose DEA model is given in (3). In contrast, under centralized management, it is assumed that the fundamental objective of the benchmarking system is improving the overall performance of the whole system. Hence, it is expected that a branch which proves inefficient under the more flexible structure of the decentralized scenario should never have a positive contribution to the overall efficiency of the system where a more restricted centralized framework is applied. This once again demonstrates that the way of computing the contribution scores is problematic in the approach proposed by Varmaz et al (2013).

The last three columns in Table 6 summarize the results of our alternative approach under centralized management. The column represented by θ_p^{CM} reports the efficiency scores computed by the centralized DEA model in (9). Three branches (1, 7 and 12) are efficient. These three branches are also super-efficient, as Eff_o^1, Eff_o^7 and Eff_o^{12} capture values greater than 0.835 – the overall efficiency score of the entire system. Therefore, their contributions to the overall efficiency

should also be positive. The last column, represented by $\tilde{\theta}_p^{CM}$, confirms the positive contribution of these branches to the overall efficiency of the system.

Consider now branches #2 and #14. Unlike in the approach of Varmaz at al. (2013), in our approach they receive a negative contribution score of 0.978 and 0.835, respectively. In particular, the individual contribution score of these branches under centralized management never exceeds the super-efficiency score under decentralized management, i.e. $\tilde{\theta}_p^{CM} \leq \theta_p^s$ (see Theorem 2). This verifies numerically that our proposed approach overcomes the problem identified in the approach of Varmaz et al. (2013).

Their approach can also lead to a set of unexpected compensation plans, as Table 7 reveals. It shows plans which have been determined by utilizing formula (2) with the three measures of efficiency θ_p^s (normal super-efficiency of p), $\tilde{\theta}_p$ (Varmaz et al. 2013 approach) and $\tilde{\theta}_p^{CM}$ (our approach). It should be noted that c_p represents the aggregate operating cost level of each branch, computed as the sum of the two inputs. For illustrative purposes, ρ_p in (2) has been set at 0.7 for all branches.

Table 7. Alternative compensation plans

Unit	θ_p^s	$\tilde{\theta}_p$	$\tilde{\theta}_p^{CM}$	c_p	b_p^s	$\tilde{b}_p$	$\tilde{b}_p^{CM}$
1	Na.	1.033	1.014	4301	Na.	4400.18	4343.17
2	0.981	1.022	0.978	2755	2717.54	2798.01	2711.94
3	0.818	0.991	0.753	2073	1808.63	2059.65	1714.77
4	1.075	0.998	0.755	558	587.34	557.13	462.15
5	0.699	0.988	0.674	1460	1152.09	1447.40	1126.59
6	1.242	0.987	0.735	2578	3015.08	2554.14	2099.58
7	1.182	0.995	1.012	656	739.59	653.67	661.43
8	0.837	0.996	0.786	1492	1321.79	1487.95	1268.56
9	0.892	0.993	0.643	750	693.22	746.27	562.56
10	0.831	0.992	0.669	964	850.11	958.67	740.89
11	0.923	0.997	0.817	2526	2390.71	2521.43	2201.98
12	1.176	0.982	1.027	725	814.26	716.03	738.94
13	0.807	0.993	0.643	670	579.49	666.94	502.43
14	0.864	1.000	0.835	1795	1623.78	1795.00	1587.73
15	1.037	0.993	0.590	537	551.03	534.55	382.94
16	1.262	1.002	0.924	537	635.52	537.93	508.35

In Table 7, compensation plans based upon θ_p^s, $\tilde{\theta}_p$ and $\tilde{\theta}_p^{CM}$ are shown by b_p^s, $\tilde{b}_p$ and $\tilde{b}_p^{CM}$, respectively. There exist compensation plans, determined by the approach proposed by Varmaz et al. (2013), which are greater than those computed within the decentralized framework, i.e. $\tilde{b}_p > b_p^s$. Taking into account the inefficient branches, which have been recognized within the decentralized framework (i.e. branches 2, 3, 5, 8, 9, 10, 11, 13 and 14), the results by Varmaz et al. (2013)

reflected in $\tilde{b}_p$ are counter-intuitive. Although these branches are inefficient in a decentralized framework, the centralized compensation plans reflected in $\tilde{b}_p$ are greater than those reflected in b_p^s under decentralized management. Thus, the Varmaz et al (2013) approach fails to provide a set of intuitive compensation plans which would incentivise such inefficient branches to contribute as much as possible to the overall performance under centralized management. By contrast, the compensation plans in our approach never exceed those determined in the decentralized framework, i.e. we always have $b_p^s \geq \tilde{b}_p^{CM}$. This is in line with the expectation that under central management, branches should receive an incentive to always operate in the best interest of the entire organization in order to improve the overall performance of the system. This could lead to resources being diverted from some branches to others for the benefit of the system as a whole, and hence $b_p^s \geq \tilde{b}_p^{CM}$.

One can alternatively apply the extended approach in Section 5 to determine a joint-contribution value of highly efficient units. Let us assume that r is set to be 8. On this basis, branches with a contribution score higher than or equal 0.920 are included in S_{92}, i.e. branches 1, 2, 7, 12 and 16. For illustrative purposes and as the current number of branches in S_{92} exceeds 30% of the full set of branches, the procedure for identifying additional potentially masked super-efficient units is not applied further, i.e. FS_{92} includes the branches 1, 2, 7, 12 and 16. According to the proposed procedure, the remaining branches – with a contribution value less than 0.920 – receive automatically the same contribution values as before. Hence, Table 8 only contains the contribution scores and the corresponding compensation plans of the branches in S_{92}.

Table 8. Results of joint contributions and the compensation plans for units in S_{92}

Unit	θ_p^{CM}	$\tilde{\theta}_p^{CM}$	$\tilde{\theta}_p^{CM}$ (joint)	c_p	$\tilde{b}_p^{CM}$	$\tilde{b}_p^{CM}$ (joint)
1	1.000	1.014	1.231	4301	4343.17	4997.56
2	0.978	0.978	1.204	2755	2711.94	3148.16
7	1.000	1.012	1.231	656	661.43	762.24
12	1.000	1.027	1.231	725	738.94	842.42
16	0.924	0.924	1.137	537	508.35	588.69

In Table 8, the new contribution scores are reported in the fourth column with a label of "joint". As can be seen in this table, all branches now receive a positive contribution value, cf. the third and fourth columns. The reason is that the modified model in (10) measures now the efficiency of the whole system of bank branches while its potential benchmarks – represented by the constraint (12) – do not include branches in FS_{92}, i.e. branches 1, 2, 7, 12 and 16. In other words, in measuring Eff_o^p

for each one of the branches in FS$_{92}$, the performance of the whole system is compared to another system which excludes jointly the units with a θ_p^{CM} efficiency value greater than or equal to 92%.

As a consequence, the contribution values of branches #1, #7 and #12 are all now higher than before, cf. the third and fourth columns in Table 8. As can be seen in the sixth and seventh columns in Table 8, this also leads to a higher level of incentives in terms of the award of a budget in excess of what these branches have spent, i.e. $c_p \leq \tilde{b}_p^{CM} \leq \tilde{b}_p^{CM}(\text{joint})$. Furthermore, branches #2 and #16 now show a positive contribution score as their super-efficiency scores in this procedure have been deemed to be significant. These results also affected the compensation plan of the branches reported in the last column of Table 8. The costs of branches #2 and #16 are now overcompensated, i.e. $c_2 \leq \tilde{b}_2^{CM}(\text{joint})$ and $c_{14} \leq \tilde{b}_{14}^{CM}(\text{joint})$. Hence, they receive high-powered incentives to operate in the best interest of the entire system. The higher compensations in the rightmost column of Table 8 (column labelled "joint") compared to the compensations for those units under the scheme reflected in Table 7 are intuitively acceptable, as the "joint" compensation levels of the jointly super-efficient units provide a better measure of the true level of super-efficiency of the units concerned.

7. Conclusions and Outlook on Future Research

Recently, Varmaz et al. (2013) proposed a framework of how to incentivise to better performance the units of multi-unit organisations operating under centralized management. Their approach is centred on computing a form of super-efficiency for each unit applying a variant of the centralized resource allocation DEA (CRA-DEA) model of Lozano and Villa (2004). They use this measure of super-efficiency within the framework of the DEA-based formula proposed by Bogetoft (1997) as a basis for computing resource allocations to units to incentivize them to better performance. Here, we revisit their approach and show that it can lead to counter-intuitive results, incompatible with incentivizing units to improve their performance. The key pitfall in their approach is that the so-called super-efficiency measure does not capture appropriately the impact of a unit on the system as a whole because the system is not defined in a stable manner.

In order to remedy this deficiency, we adapt and extend their approach in the following way. Firstly, we capture the impact of a unit on the system of units by defining the latter in a stable manner across alternative DEA assessments. Then, we measure the impact on the system of units by reference to how the efficient frontier of the system of units varies with and without each unit under consideration. This leads to measures of efficiency or super-efficiency, as the case may be. The measures of efficiency are then used to encourage units to cover a proportion of their shortfall in

efficiency from 100%. If they are super-efficient, an incentive component is identified in proportion to the degree by which they can impact the performance of the overall system of units when acting as a benchmark for other units in the system. Finally, we extend further the approach to deal with the possibility that in some data sets one unit can mask the impact of another unit as a benchmark. This can happen when the two units are virtually identical in performance and sit on or are close to the efficient frontier. We suggest a practical approach for overcoming this problem so that all units that are similarly good performers are rewarded appropriately to incentivize them to better performance.

In order to numerically illustrate the features of our proposed approach and to show its practical relevance, we analyse the performance of 16 branches of a small German retail bank. We demonstrate how our approach leads to the incentive component for super-efficient branches and how it copes with cases where certain units mask the super-efficiency of others, allowing for an appropriate incentivisation of such units. Our analysis demonstrates that our approach overcomes the problems the approach of Varmaz et al. (2013) can cause in incentivising units of centralised multi-unit organisations.

Our work draws upon and adds to the broad field of incentive regulation when it is seen as a principal-agent problem where there is asymmetry of information. In our context, the principal is the central management, while the agents are the units of a multi-unit organisation. Respective research into incentivizing operating units of centrally managed multi-unit organisations is at a relatively early stage. As organisations are in general subject to on-going changes, the time dimension should be addressed in the context of incentivisation of units. Furthermore, the incorporation of value judgements as to the direction of improvement of the system as a whole has not been addressed to our knowledge. Depending on the importance attached to improving individual inputs and outputs at system level, there could be varying implications for the incentives given to individual units to improve performance. Thus, both the use of panel data and value judgements in the context of incentivising units under centralised management are interesting future research avenues.

Acknowledgement

References

Abri, A. G., Jahanshahloo, G. R., Lotfi, F. H., Shoja, N., Jelodar, M. F. (2013). A new method for ranking non-extreme efficient units in Data Envelopment Analysis. *Optim. Lett*. 7(2), 309-324.

Afsharian, M., Ahn, H. (2014). The Luenberger indicator and directions of measurement: A bottoms-up approach with an empirical illustration to German savings banks. *Int. J. Prod. Res*. 52(20), 6216-6233.

Agrell, P. J., Bogetoft, P., Tind, J. (2002). Incentive plans for productive efficiency, innovation and learning. *Int. J. Prod. Econ*. 78(1), 1-11.

Agrell, P. J., Bogetoft, P., Tind, J. (2005). DEA and dynamic yardstick competition in Scandinavian electricity distribution. *J. Prod. Anal*. 23(2), 173-201.

Ahn, H., Neumann, L., Vazquez Novoa, N. (2012). Measuring the relative balance of DMUs. *Eur. J. Oper. Res.* 221(2), 417-423.

Andersen, P., Petersen, N. C. (1993). A procedure for ranking efficient units in Data Envelopment Analysis. *Manag. Sci.* 39(10), 1261-1264.

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

Athanassopoulos, A. D. (1995). Goal programming & Data Envelopment Analysis (GoDEA) for target-based multi-level planning: Allocating central grants to the Greek local authorities. *Eur. J. Oper. Res.* 87(3), 535-550.

Avkiran, N. K. (2009). Opening the black box of efficiency analysis: An illustration with UAE banks. *Omega*. 37(4), 930-941.

Banker, R. D., Chang, H. (2006). The super-efficiency procedure for outlier identification, not for ranking efficient units. *Eur. J. Oper. Res.* 175(2), 1311-1320.

Banker, R. D., Charnes, A., Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in Data Envelopment Analysis. *Manag. Sci.* 30(9), 1078-1092.

Bogetoft, P. (1994). Incentive efficient production frontiers: An agency perspective on DEA. *Manag. Sci*. 40(8), 959-968.

Bogetoft, P. (1997). DEA-based yardstick competition: The optimality of best practice regulation. *Ann. Oper. Res*. 73, 277-298.

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

Boussofiane, A., Dyson, R. G., Thanassoulis, E. (1991). Applied data envelopment analysis. *Eur. J. Oper. Res.* 52(1), 1-15.

Chen, Y. (2004). Ranking efficient units in DEA. *Omega.* 32(3), 213-219.

Chen, Y. (2005). Measuring super-efficiency in DEA in the presence of infeasibility. *Eur. J. Oper. Res.* 161(2), 545-551.

Chen, Y., Liang, L. (2011). Super-efficiency DEA in the presence of infeasibility: One model approach. *Eur. J. Oper. Res.* 213(1), 359-360.

Chen, C. M., Zhu, J. (2011). Efficient resource allocation via efficiency bootstraps: An application to R&D project budgeting. *Oper. Res*. 59(3), 729-741.

Cook, W. D., Zhu, J. (2007). Within-group common weights in DEA: An analysis of power plant efficiency. *Eur. J. Oper. Res.* 178(1), 207-216.

Fang, L. (2013). A generalized DEA model for centralized resource allocation. *Eur. J. Oper. Res*. 228(2), 405-412.

Kao, C., Hung, H. T. (2005). Data Envelopment Analysis with common weights: The compromise solution approach. *J. Oper. Res. Soc*. 56(10), 1196-1203.

Lee, H. S., Chu, C. W., Zhu, J. (2011). Super-efficiency DEA in the presence of infeasibility. *Eur. J. Oper. Res.* 212(1), 141-147.

Liu, J., Tone, K. (2008). A multistage method to measure efficiency and its application to Japanese banking industry. *Socio-Econ. Plan. Sci*. 42(2), 75-91.

Lozano, S., Villa, G. (2004). Centralized resource allocation using Data Envelopment Analysis. *J. Prod. Anal*. 22(1-2), 143-161.

Mar-Molinero, C., Prior, D., Segovia, M. M., Portillo, F. (2014). On centralized resource utilization and its reallocation by using DEA. *Ann. Oper. Res*. 221(1), 273-283.

Nesterenko, V., Zelenyuk, V. (2007). Measuring potential gains from reallocation of resources. *J. Prod. Anal*. 28(1-2), 107-116.

Roll, Y., Cook, W. D., Golany, B. (1991). Controlling factor weights in Data Envelopment Analysis. *IIE Trans*. 23(1) 2-9.

Roll, Y., Golany, B. (1993). Alternate methods of treating factor weights in DEA. *Omega*. 21(1), 99-109.

Seiford, L. M., Zhu, J. (1998). An acceptance system decision rule with data envelopment analysis. *Comp. Oper. Res*. 25(4), 329-332.

Seiford, L. M., Zhu, J. (1999). Infeasibility of super efficiency Data Envelopment Analysis models. *INFOR*. 37(2), 174-187.

Shleifer, A. (1985). A theory of yardstick competition. *RAND J. Econ*. 319-327.

Soares de Mello, J. C., Meza, L. A., da Silveira, J. Q., Gomes, E. G. (2013). About negative efficiencies in Cross Evaluation BCC input oriented models. *Eur. J. Oper. Res*. 229(3), 732-737.

Thanassoulis, E. (1999). Setting achievement targets for school children. *Educ. Econ*. 7(2), 101-119.

Thanassoulis, E. (2001). *Introduction to the theory and application of Data Envelopment Analysis*. Springer, New York.

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

Zhu, J. (2001). Super-efficiency and DEA sensitivity analysis. *Eur. J. Oper. Res*. 129(2), 443-455.

Zohrehbandian, M., Makui, A., Alinezhad, A. (2010). A compromise solution approach for finding common weights in DEA: An improvement to Kao and Hung's approach. *J. Oper. Res. Soc.* 61(4), 604-610.

Appendix

Theorem 2. *It is the case that $\tilde{\theta}_p^{CM} \leq \theta_p^s$ where θ_p^s is the super-efficiency of unit p as derived using the Andersen and Petersen 1993 approach.*

Proof. In formula (11), we define $\tilde{\theta}_p^{CM} = (Eff_o^p / Eff_o) \times \theta_p^{CM}$ where Eff_o, θ_p^{CM} and Eff_o^p are computed by models (5), (9) and (10), respectively.

In Scenario 1, when $\theta_p^{CM} < 1$ we have $\tilde{\theta}_p^{CM} = \theta_p^{CM}$ due to the fact that $Eff_o = Eff_o^p$. Theorem 1 implies that $\theta_p^{CM} \leq \theta_p$ and since the efficiency θ_p of DMU p cannot exceed its super-efficiency θ_p^s we have $\tilde{\theta}_p^{CM} \leq \theta_p^s$.

In Scenario 2.1, we have $\theta_p^{CM} = 1$ and $Eff_o = Eff_o^p$. Thus following the same reasoning as for Scenario 1 above we conclude $\tilde{\theta}_p^{CM} \leq \theta_p^s$.

In Scenario 2.2, we have $\theta_p^{CM} = 1$ and $Eff_o < Eff_o^p$. This implies that unit p is one of the benchmarks for the "aggregate unit" to which we refer as unit G. Let G^* represent the projection of G on the efficient boundary *before* the boundary was revised to exclude DMU p. Two possibilities arise. Either G^* is coincident in location with p or G^* is in the interior of the efficient facet which was created as a linear combination of DMUs including DMU p.

In the rather unlikely situation that G^* is coincident with p we have $\tilde{\theta}_p^{CM} = Eff_o^p / Eff_o$ since $\theta_p^{CM} = 1$. Note also that we have $Eff_o = Eff_o^p / \theta_p^s(G^*)$. Hence, this implies that $\tilde{\theta}_p^{CM} = \theta_p^s(G^*)$. We also have $\theta_p^s = \theta_p^s(G^*)$ due to coincidence of p and G^*. Therefore, we conclude that $\tilde{\theta}_p^{CM} \leq \theta_p^s$.

Finally, when G^* is in the interior of the efficient facet which was created as a linear combination of DMUs including DMU p, the super-efficiency of an extreme point p of that facet, shown by θ_p^s, will never be less than that of an interior point of that facet G^* (see Abri et al. 2013). Thus, we have $\theta_p^s(G^*) \leq \theta_p^s$. As above we have $\tilde{\theta}_p^{CM} = Eff_o^p / Eff_o$ since $\theta_p^{CM} = 1$ and $Eff_o = Eff_o^p / \theta_p^s(G^*)$.