DEA Models with Production Trade-offs and Weight Restrictions

Victor V. Podinovski

Warwick Business School, University of Warwick, Coventry CV4 7AL, United Kingdom
e-mail: v.podinovski@warwick.ac.uk

Abstract

There is a large literature on the use of weight restrictions in multiplier DEA models. In this chapter we provide an alternative view of this subject from the perspective of dual envelopment DEA models in which weight restrictions can be interpreted as production trade-offs. The notion of production trade-offs allows us to state assumptions that certain simultaneous changes to the inputs and outputs are technologically possible in the production process. The incorporation of production trade-offs in the envelopment DEA model, or the corresponding weight restrictions in the multiplier model, leads to a meaningful expansion of the model of production technology. The efficiency measures in DEA models with production trade-offs retain their traditional meaning as the ultimate and technologically realistic improvement factors. This overcomes one of the known drawbacks of weight restrictions assessed using other methods. In this chapter we discuss the assessment of production trade-offs, provide the corresponding theoretical developments and suggest computational methods suitable for the solution of the resulting DEA models.

Keywords: Data envelopment analysis, production trade-offs, weight restrictions

1 Introduction

The conventional variable and constant returns-to-scale (VRS and CRS) DEA models can each be stated as two mutually dual linear programs: as an envelopment or multiplier model (Charnes et al. 1978; Banker et al. 1984). The envelopment model is based on an explicit representation of the production technology. The efficiency of decision making units (DMUs) in this model is obtained by their input or output radial projection on the boundary of the technology. The dual multiplier model is stated in terms of variable vectors of input and output weights. This model assesses the efficiency of DMUs in terms of the ratio of their aggregated weighted outputs to aggregated weighted inputs, in relation to similar ratios calculated for all observed DMUs.

One common modification of the multiplier model is based on the use of weight restrictions – the incorporation among its constraints of additional inequalities on the input and output weights (Thanassoulis et al. 2008; Cooper et al. 2011a). Weight restrictions are attractive because of their apparent managerial meaning and also because their use can significantly improve the efficiency discrimination of DEA models (Allen et al. 1997; Thanassoulis et al. 2004).

A well-known drawback of weight restrictions arises from the fact that their use in the multiplier model implicitly changes the model of production technology in the envelopment form (Allen et al. 1997). Specifically, weight restrictions enlarge the model of technology and generally shift the efficient frontier to a more demanding level, as illustrated by Roll et al. (1991). An obvious problem with this is that the efficient projections of inefficient DMUs located on the expanded frontier may not be producible (technologically realistic). Furthermore, the traditional meaning if efficiency as the ultimate and *technologically feasible* improvement factor generally becomes unsubstantiated (Podinovski 2004a; Førsund 2013).

The purpose of this chapter is to describe an approach to the construction of weight restrictions that *by definition* does not have the above drawback. The idea is to consider the dual forms of weight restrictions induced in the envelopment models. These are additional terms that are simultaneously added to, or subtracted from, the inputs and outputs of the units in the production technology. Following Podinovski (2004a), we refer to these terms as *production trade-offs*.

Weights restrictions and production trade-offs are mathematically equivalent. From the practical point of view they may, however, be regarded as different tools. While the terminology of weight restrictions is a natural language for the elicitation and communication of value judgements, the notion of production trade-offs makes us think in terms of production technology and possible substitutions between its inputs and outputs.

Production trade-offs do not generally follow from the data – instead, they are additional assumptions that we (or experts) are willing to make about the production technology: that a certain simultaneous change (substitution) of inputs and outputs is technologically possible, at all units.

In this respect production trade-offs should not be confused with *marginal rates* of transformation and substitution between the inputs and outputs. The latter represent the slopes of the supporting hyperplanes to the technology and are generally different at different boundary units. Changing the inputs and outputs of a boundary unit in the proportions based on the marginal rates (calculated at this unit) would keep the resulting unit on the supporting hyperplane – this does not mean that the resulting unit is producible. In other words, marginal rates represent the movements (changes to inputs and outputs) that are tangent to the technology and are not supposed to result in producible units. In contrast, production trade-offs represent movements that are not necessarily tangent to the boundary of the true technology (that we are attempting to model), but are assumed to keep the resulting units technologically possible.

The use of production trade-offs for the construction of weight restrictions has been illustrated in different contexts. These include the assessment of efficiency of university departments (Podinovski 2007a), secondary schools (Khalili et al. 2010), primary health care providers (Amado and Santos 2009), primary diabetes care providers (Amado and Dyson 2009), electricity distributors (Santos et al. 2011) and agricultural farms (Atici 2012). The following are a few examples of production trade-offs employed in the above studies.

1. *Primary health care provision:* the hospital outputs should not deteriorate if the number of nurses is reduced by 1 and the number of doctors is increased by 1 (Amado and Santos 2009). This corresponds to the weight restriction stating that the weight attached to the number of doctors is at least as large as the weight attached to the number of nurses.
2. *Electricity distribution:* a distribution utility should be able to increase the delivery of electricity by at least 40 KWh per Euro of increase of operating expenses – the latter is chosen as a representative measure for all distribution costs (Santos et al. 2011). This implies that the weight attached to operating expenses (in Euros) is greater than or equal to 40 times the weight attached to the number of KWh delivered.
3. *Agricultural farms:* the resources required for the production of 1 tonne of wheat are sufficient for the production of at least 0.75 tonnes of barley, at any farm in the given

region (Atici 2012). This implies that the weight attached to wheat is greater than or equal to 0.75 times the weight attached to barley.

Production trade-offs have exactly the same effect on the model of technology as weight restrictions: the technology expands but, in contrast with the latter case, in *a controlled way* that we explicitly assume to be technologically possible. Because the expanded technology and, therefore, its efficient frontier are realistic in the production sense, this further implies that the radial targets are producible and the efficiency measures retain their conventional technological meaning as possible improvement factors.

The use of production trade-offs overcomes the known drawbacks of weight restrictions not because they are different: as noted, both are equivalent concepts. The advantage of trade-offs is that their assessment explicitly refers to the technology and requires our judgement to be stated in the language of possible changes to inputs and outputs. The assessed trade-offs can be incorporated either in the envelopment model (which currently requires the use of a general linear optimiser), or as equivalent weight restrictions in the multiplier model (which can be performed in most current DEA solvers). In the latter case, the weight restrictions do not have the above known general drawbacks because they are constructed by transformation of production trade-offs. We call this method *the trade-off approach* to the construction of weight restrictions.

It is worth mentioning that several earlier studies came close to the notion of production trade-offs. Charnes et al. (1989), Roll et al. (1991) and Halme and Korhonen (2000) show that the incorporation of weight restrictions in multiplier models induce dual terms that change the technology but do not explore this relation as a basis for the assessment of weight restrictions that have a production meaning.

The assessment of weight restrictions in some earlier applications of DEA can also be viewed as being *implicitly* based on (or consistent with) the idea of production trade-offs. Dyson and Thanassoulis (1988) consider a DEA model with a single input. In this study the lower bound on each output weight is related to the minimum amount of the input required per unit of the output. This is essentially a statement of a production trade-off, although in a specific DEA model that cannot be easily generalised to the case of multiple outputs. In the assessment of bank branch performance, Schaffnit et al. (1997) and Cook and Zhu (2008) incorporate limits on the ratios of the weights of different transaction and maintenance activities. Such limits are based on the lower and upper bounds on the amounts of time that such activities require and effectively express production trade-offs between the activities.

2 Production Trade-offs

Following Podinovski (2004a), in this section we introduce production trade-offs as the dual forms of weight restrictions. It is also straightforward to introduce production trade-offs independently and establish their dual relationship to weight restrictions afterwards. We prefer the former approach because it builds up on the already well-established concept of weight restrictions in the DEA literature.

Consider technology $\mathcal{T} \subset \mathbb{R}_+^m \times \mathbb{R}_+^s$ with $m \geq 1$ inputs and $s \geq 1$ outputs. The elements of $\mathcal{T}$ are DMUs stated as the pairs (X, Y), where X and Y are the input and output vectors, respectively. Let $J = \{1,...,n\}$ be the set of observed DMUs. Each observed DMU can also be stated as (X_j, Y_j), where $j \in J$. Denote (X_o, Y_o) the unit in $\mathcal{T}$ whose efficiency is being assessed.

In order for the DEA models to be well-defined and avoid the consideration of special cases, we make the following standard data assumption: at least one input and one output of each observed DMU is strictly positive. We also assume that every output $r = 1,...,s$ is strictly positive for at least one observed DMU $j_1 \in J$, and every input $i = 1,...,m$ is strictly positive for at least one observed DMU $j_2 \in J$.

Let $v \in \mathbb{R}_+^m$ and $u \in \mathbb{R}_+^s$ be, respectively, the vectors of input and output weights used in the multiplier DEA models. Consider the following $K \geq 1$ homogeneous weight restrictions:

$$u^\top Q_t - v^\top P_t \leq 0, \qquad t = 1,...,K, \tag{1}$$

where $P_t \in \mathbb{R}^m$ and $Q_t \in \mathbb{R}^s$ are some constant vectors, for all t, and symbol $^\top$ denotes transposition. Components of vectors P_t and Q_t can be positive, negative or zero. If, for some t, both $P_t \neq 0$ and $Q_t \neq 0$, the corresponding weight restriction t in (1) is called *linked* and is often referred to as Assurance Region II (Thompson et al. 1990). Otherwise, the weight restriction is *not linked* and is termed Assurance Region I, or polyhedral cone ratio (Charnes et al. 1989, 1990).

Suppose we wish to assess the efficiency of some DMU (X_o, Y_o) using the multiplier model with weight restrictions (1). To be specific, we consider the case of CRS first, and comment on the case of VRS afterwards.

The input radial efficiency of DMU (X_o, Y_o) is obtained as the optimal value θ^* in the following multiplier model that incorporates weight restrictions (1):

Model $\mathbb{M}^1_{CRS}$*:*

$$\theta^* = \max \quad u^\top Y_o, \tag{2}$$

$$\begin{aligned} \text{subject to} \quad & v^\top X_o = 1, \\ & u^\top Y_j - v^\top X_j \leq 0, \qquad j = 1,...,n, \\ & u^\top Q_t - v^\top P_t \leq 0, \qquad t = 1,...,K, \\ & u, v \geq 0. \end{aligned}$$

(In model (2) and below, the vector inequalities ≤ and ≥ mean that the corresponding inequality is true for each component.)

Note that, although model (2) maximises the aggregated output $u^\top Y_o$ of DMU (X_o, Y_o), its dual envelopment form (4) presented below projects the latter unit on the boundary of the technology by the radial contraction of the input vector X_o. This explains why model (2) and its dual are conventionally referred to as input-minimisation, or input-oriented, models (Cooper et al. 2011b).

Similarly, the output radial efficiency of DMU (X_o, Y_o) is equal to the inverse $1/\eta^*$ of the optimal value η^* in the following output-maximisation (or output-oriented) multiplier model:

Model $\mathbb{M}^2_{CRS}$*:*

$$\eta^* = \min \quad v^\top X_o, \tag{3}$$

$$\begin{aligned} \text{subject to} \quad & u^\top Y_o = 1, \\ & u^\top Y_j - v^\top X_j \leq 0, \qquad j = 1,...,n, \\ & u^\top Q_t - v^\top P_t \leq 0, \qquad t = 1,...,K, \\ & u, v \geq 0. \end{aligned}$$

The notion of production trade-offs and their relation to weight restrictions becomes apparent when we consider the envelopment models dual to (2) and (3). Using vectors $\lambda = (\lambda_1,...,\lambda_n)$ and $\pi = (\pi_1,...,\pi_K)$, the dual to (2) can be stated as follows:

Model $\mathbb{E}^1_{CRS}$:

$$\theta^* = \min \quad \theta, \tag{4}$$

subject to $\sum_{j=1}^{n} \lambda_j Y_j + \sum_{t=1}^{K} \pi_t Q_t \geq Y_o,$

$$\sum_{j=1}^{n} \lambda_j X_j + \sum_{t=1}^{K} \pi_t P_t \leq \theta X_o,$$

$$\lambda \geq 0,\ \pi \geq 0,\ \theta \text{ sign free.}$$

Similarly, the dual to (3) is the envelopment model

Model $\mathbb{E}^2_{CRS}$:

$$\eta^* = \max \quad \eta, \tag{5}$$

subject to $\sum_{j=1}^{n} \lambda_j Y_j + \sum_{t=1}^{K} \pi_t Q_t \geq \eta Y_o,$

$$\sum_{j=1}^{n} \lambda_j X_j + \sum_{t=1}^{K} \pi_t P_t \leq X_o,$$

$$\lambda \geq 0,\ \pi \geq 0,\ \eta \ \text{ sign free.}$$

In both envelopment models (4) and (5) the first group of terms on the left-hand side of their constraints defines a composite unit $(\bar{X}\lambda, \bar{Y}\lambda)$ in the conventional CRS technology. This unit is further modified by the pairs of vectors

$$(P_t, Q_t), \qquad t = 1,\ldots,K, \tag{6}$$

used in proportions $\pi_t \geq 0$. In particular, vector P_t represents changes to the inputs, and vector Q_t shows changes to the outputs. Therefore, each pair (P_t, Q_t) in (6) can be referred to as *a production trade-off*.

It is clear that for some weight restrictions (1) the corresponding production trade-offs (6) may not represent a technologically possible substitution between the inputs and outputs. In this case the unit obtained on the left-hand side of models (4) and (5) is generally not producible. An obvious way to overcome this problem is to construct technologically realistic trade-offs (6) in the first place. The efficiency of DMU (X_o, Y_o) can then be assessed by solving either the envelopment models (4) and (5) or their dual multiplier models (2) and (3). In the latter case, the trade-offs (6) should be converted to weight restrictions (1).

The above process describes *the trade-off approach* to the construction of weight restrictions. Its idea is that the weight restrictions (1) are assessed in the dual envelopment space where they take on the form of production trade-offs (6). The latter are essentially additional production assumptions based on our understanding of the technology. Examples illustrating the trade-off approach are discussed in Section 3 below.

In the case of VRS, the dual relationship between weight restrictions (1) and production trade-offs (6) is the same as above. The VRS analogues of the CRS envelopment models are the programs (4) and (5) with the additional normalising condition

$$\sum_{j=1}^{n} \lambda_j = 1. \tag{7}$$

Below we denote to the resulting envelopment VRS models as $\mathbb{E}_{VRS}^1$ and $\mathbb{E}_{VRS}^2$, respectively. The corresponding dual multiplier models are referred to as $\mathbb{M}_{VRS}^1$ and $\mathbb{M}_{VRS}^2$. Both VRS multiplier models utilise an additional sign free variable u_0 dual to equality (7).

Remark 1. The case of non-homogeneous weight restrictions is considered in Podinovski (2004a; 2005). Such restrictions have a non-zero constant on the right-hand side of inequalities (1), an example of which is absolute weight bounds (Dyson and Thanassoulis 1988). Non-homogeneous weight restrictions can also be related to production trade-offs in the envelopment model, but formula (6) is no longer valid. The exact trade-off induced by a non-homogeneous weight restriction depends on the DMU (X_o, Y_o) under the assessment and the orientation (input minimisation or output maximisation) of the model. This complicates the assessment of non-homogeneous weight restrictions and makes them less attractive in practical applications.

A further difficulty arising in DEA models with non-homogeneous weight restrictions is that the managerial meaning of the resulting efficiency obtained via the multiplier DEA model may be unclear. In particular, the optimal input and output weights in the resulting models do not generally represent the assessed DMU in the best light compared to the other DMUs (Podinovski and Athanassopoulos 1998; Podinovski 1999; 2004b).

3 Illustrative Example

Below we consider an example that illustrates the use of production trade-offs in the assessment of efficiency of academic departments from different universities using a hypothetical data set. The departments are assumed to be from the same academic area (e.g., economics). The choice of inputs and outputs in this example is the same as in Podinovski (2007a) but the data set is different.

Table 1 shows seven hypothetical university departments denoted D1, D2, …, D7. The two inputs include full academic staff and research staff. The three outputs include undergraduate students, master (postgraduate) students and academic publications.

To be specific, we consider the case of output radial efficiency. Using the two conventional CRS and VRS output-maximisation DEA models, we obtain the efficiency scores as shown in the second left columns in Tables 2 and 3 titled "CRS" and "VRS", respectively. It is not surprising that, given the small set of observed DMUs, the efficiency discrimination is low: in the case of CRS only two departments are inefficient, and in the case of VRS only one is inefficient.

Table 4 shows the optimal input and output weights obtained in the standard CRS model. The weights u_1, u_2 and u_3 correspond to the three outputs: undergraduate students, master students and publications, respectively. The weights v_1 and v_2 correspond to the two inputs: academic and research staff, respectively. Although optimal weights are generally not unique, the weights in Table 4 are consistent with the known drawback of conventional DEA models: the complete flexibility of weights often results in zero weights attached to some of the inputs and outputs. These represent the areas in which the DMU under the assessment is relatively weak (Thanassoulis et al. 1987; Dyson and Thanassoulis 1988).

For example, department D4 has a relatively low number of students per member of staff but the highest number of publications per staff. This is reflected in the optimal weights attached to these outputs: both undergraduate and master students have a zero weight attached to them. This implies that the DEA model used for the assessment of department D4 effectively ignores the first two outputs. Exactly the same efficiency score for department D4 is obtained if we remove the two types of student from model specification and assess the efficiency of D4 based on the two inputs and publications only.

Let us show that both the CRS and VRS DEA models can be improved using simple production trade-offs.

3.1. Undergraduate and Master Students

We start by comparing the resources (academic staff) that are used by the departments for the teaching of undergraduate and master students.

Assumption 1. The teaching of one undergraduate student does not require more resources (academic staff time) than the teaching of one master student.

We can restate the above assumption as the following trade-off that all departments should accept:

$$P_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix},\ Q_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}. \tag{8}$$

The meaning of the above trade-off is straightforward: it is possible to replace one master student (the value –1 in the second component of vector Q_1) by one undergraduate student (the value 1 in the first component of vector Q_1). For this replacement, no change of the inputs (resources) is needed: vector P_1 is a zero vector. There should also be no change to the third output (publications): the third component of vector Q_1 is zero.

Production trade-off (8) can be restated as a weight restriction using formula (1):

$$u_1 - u_2 \leq 0. \tag{9}$$

This inequality implies that the weight attached to master students cannot be less than the weight attached to undergraduate students. The same weight restriction may possibly be obtained by a value judgement but it is the original production trade-off (8) that makes this weight restriction meaningful in the production sense.

Assumption 2. The teaching of a master student may require more resources than an undergraduate student, however, by no more than a factor of 3.

Note that the above assumption is not a precise measure of the relative amount of resources required by the two types of output, and it should not be for two reasons. First, the estimates of this ratio may vary depending on the methodology used for its calculation even for one particular department. Second, even if the precise ratio were possible to assess, this would most likely vary between the departments. Because of these uncertainties, Assumption 2 is supposed to be a safe conservative estimate (an upper bound of different possible estimates) that all departments should agree on.

We state Assumption 2 as the following production trade-off:

$$P_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix},\ Q_2 = \begin{pmatrix} -3 \\ 1 \\ 0 \end{pmatrix}. \tag{10}$$

The above trade-off means that no extra resources should be claimed (P_2 is a zero vector) and there should be no detriment to the publications if the number of undergraduate students is reduced by 3 and the number of master students is increased by 1. Using formula (1), production trade-off (10) is restated as the weight restriction

$$-3u_1 + u_2 \leq 0 . \tag{11}$$

According to (11), the weight attached to master students cannot be more than 3 times larger than the weight attached to undergraduate students. Note that the factor 3 does not reflect the perceived importance of master students compared to undergraduates, as both outputs may be deemed equally important for the departments or the decision maker who is assessing their efficiency. The factor 3 is obtained as (the upper bound on) the ratio of the resources that these two outputs require. This should be acceptable to all departments.

3.2. Research Staff and Publications

Consider the role of research staff in producing academic publications. Because the rate of publications may vary between different departments and individual researchers, the following two assumptions are intended to be sufficiently conservative.

Assumption 3. Each researcher should be able to publish at least one paper in two years.

The above statement can be stated as the following production trade-off:

$$P_3 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \; Q_3 = \begin{pmatrix} 0 \\ 0 \\ 0.5 \end{pmatrix}. \tag{12}$$

This trade-off implies that if the number of researchers is increased by 1, it should be possible to increase the number of papers by 0.5 per year. Equivalently, using formula (1), the above trade-off translates to the following linked weight restriction:

$$0.5u_3 - v_2 \leq 0 .$$

Assumption 4. No department can justify a reduction of the number of papers by more than 6 per year by referring to a loss of one research staff.

The number 6 in the above statement is purely speculative and is simply used as an illustration of a reasonably high research output. In real applications this can be revised either way. Assumption 4 is stated as the following production trade-off:

$$P_4 = \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \; Q_4 = \begin{pmatrix} 0 \\ 0 \\ -6 \end{pmatrix}. \tag{13}$$

Equivalently, Assumption 4 can be stated as the following weight restriction:

$$-6u_3 + v_2 \leq 0\,.$$

Production trade-offs (12) and (13) effectively specify the lower and upper bounds on the number of papers per average researcher at any department. Any publication rate below the lower bound of 0.5 is treated as evidence of inefficiency. A publication rate above the upper bound of 6 is regarded as unrealistically high.

3.3. Academic Staff and Students

There are different ways in which the link between academic staff and their outputs (students and publications) can be expressed. Below we consider two statements that link one input and two outputs simultaneously in a single trade-off.

The idea of these two assumptions is based on the common use of student-to-staff ratios at academic departments and the expectation of certain publication rates.

Assumption 5. One full academic post is a sufficient resource for the number of undergraduate students at the department to increase by 10 and the number of publications to increase by 0.5.

This assumption is stated as the following production trade-off:

$$P_5 = \begin{pmatrix} 1 \\ 0 \end{pmatrix},\ Q_5 = \begin{pmatrix} 10 \\ 0 \\ 0.5 \end{pmatrix}. \tag{14}$$

It further translates to the linked weight restriction:

$$10u_1 + 0.5u_3 - v_1 \leq 0\,. \tag{15}$$

Assumption 6. A loss of one academic post should not lead to a reduction of more than 20 undergraduate students and 5 publications per year.

This assumption is represented by the following trade-off

$$P_6 = \begin{pmatrix} -1 \\ 0 \end{pmatrix},\ Q_6 = \begin{pmatrix} -20 \\ 0 \\ -5 \end{pmatrix}, \tag{16}$$

and the following equivalent weight restriction:

$$-20u_1 - 5u_3 + v_1 \leq 0\,. \tag{17}$$

3.4. Students and Publications

Three university departments in our data set, D1, D4 and D6, can be regarded as research-intensive. They have a moderate teaching-to-staff ratio and a relatively high publication rate.

Departments D3 and D7 are focused primarily on the teaching. They have a high student-to-staff ratio and a low number of publications. Overall, this suggests that the departments in Table 1 can be viewed as having *different specialisations*.

Highly specialised DMUs are often shown as efficient by DEA models. This is because the peer groups of units to which specialised units can be compared have to show a similar specialisation, which is a limiting factor. Below we overcome the above problem by relating the "production" of students and publications by means of production trade-offs. The latter are based on the evaluation of the resources (staff time) that are needed for the generation of the two outputs.

Assumption 7. The reduction of the number of undergraduate students by 20 releases the academic staff time sufficient to write one academic paper.

As a justification of the above statement, we can think of an academic member of staff being on a one-year study leave. This involves no teaching load and an expectation of several research outputs. The reduction of undergraduate students by 20 can be approximately equated to one year of staff time, and the publication of just one paper is a conservative estimate of the publication output achievable within one year. This assumption is stated as the following production trade-off:

$$P_7 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \; Q_7 = \begin{pmatrix} -20 \\ 0 \\ 1 \end{pmatrix}. \tag{18}$$

It further translates to the weight restriction:

$$-20u_1 + u_3 \le 0 .$$

Assumption 8. The reduction of the number of publications by 5 releases the academic staff time sufficient to increase the number of undergraduate students by 20.

This assumption is stated as the following trade-off:

$$P_8 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \; Q_8 = \begin{pmatrix} 20 \\ 0 \\ -5 \end{pmatrix}. \tag{19}$$

It further translates to the weight restriction:

$$20u_1 - 0.5u_3 \le 0 .$$

Taken together, trade-offs (18) and (19) put the bounds on the ratio between the resources (staff time) required to teach undergraduate students and publish papers. Namely, the teaching of 20 undergraduate students may, depending on the department, equate to the writing of between 1 and 5 papers. If the number of students is reduced by 20, any department should be able to compensate for this by increasing the number of publications by at least 1 paper per year. If the number of students is increased by 20 (and the staff number is kept constant), this may be used to justify the reduction of publications by no more than 5 papers per year.

3.5. Computational Results

Tables 2 and 3 show the output radial efficiency of all departments in the CRS and VRS DEA models with different sets of production trade-offs. We obtained these results using a common commercial solver. Obviously, solving the envelopment and corresponding multiplier models led to the same efficiency scores.

As noted above, in these two tables, the columns titled “CRS” and “VRS” correspond to the standard DEA models without production trade-offs. Models CRS *k* and VRS *k*, where $k = 1,...,8$, incorporate all production trade-offs (P_t, Q_t), $t = 1,...,k$ stated above. For example, models CRS 1 and VRS 1 incorporate the single trade-off (P_1, Q_1) as stated in (8). Models CRS 3 and VRS 3 incorporate three trade-offs $(P_1, Q_1), (P_2, Q_2)$ and (P_3, Q_3). Models CRS 8 and VRS 8 incorporate all eight production trade-offs.

Both tables allow us to observe the gradual improvement of efficiency discrimination as additional trade-offs are progressively incorporated. The final columns CRS 8 and VRS 8 show a significant improvement over the conventional CRS and VRS models.

Table 5 shows the optimal input and output weights in model CRS 8. These weights were obtained by solving the dual multiplier model with the eight weight restrictions equivalent to the production trade-offs. In comparison to Table 4, all optimal weights in Table 5 are strictly positive.

In this respect it should be noted that in practical applications of production trade-offs the aim of making all optimal weights strictly positive may be a goal that is hard to achieve. The incorporation of realistic production trade-offs (or weight restrictions based on them) is a worthwhile improvement to the DEA model, even if this does not completely eliminate all zero weights in the optimal solution.

4. Graphical Illustrations

To illustrate the effect of production trade-offs on the technology, consider the following two examples. Both are concerned with the assessment of efficiency of university departments. Note that these departments are different from those in Table 1.

Example 1. Let units *A*, *B* and *C* shown in Figure 1 be observed departments. These departments are assumed to have the same level of a single input (staff) which is not depicted, and different levels of two outputs: undergraduate and master students. Because the input is equal, the shaded area represents both the VRS and CRS technology induced by the three units. More precisely, the shaded area is the section of either technology for the given level of input. For simplicity, we still refer to this section as the technology.

The efficient frontier of this technology is the line segment *AC*. Department *B* is located on the boundary of technology but is dominated by *A*. It is therefore only weakly efficient. The output radial efficiency of all three departments is equal to 1.

Consider production trade-off (8). (We ignore the publications and research staff that are not present in this example.) By the assumption made, this trade-off can be applied to any department. For example, starting at *A*, we can increase the number of its undergraduate students by 1 and simultaneously reduce the number of master students by 1. This procedure can be repeated multiple times. Increasing the number of undergraduate students of department *A* by 100 and reducing the number of master students also by 100, we arrive at the hypothetical department *D*. Continuing this process, we induce the straight line *AW*.

We have shown that the line *AW* consists of producible units and should therefore be regarded as part of the technology. Using the free disposability of outputs, we should also add the nonnegative area below *AW* to the technology. Note that, if we start at any other unit, e.g., at *B* or *C*, the application of trade-off (8) does not add any further new points to the technology.

The use of trade-off (8) allows us to add new units in the scenario in which the number of undergraduate students is increased. To consider the reduction of this input, we need to refer to production trade-off (10). Starting at point *A* and using the same logic as above, we move away from *A* to point *U*. All points on the line *AU* are producible because we are replacing 3 undergraduate students by 1 master student in this process, as in trade-off (10). This adds the line *AU* to the technology, along with the dominated nonnegative region below it.

Overall, the specification of two trade-offs (8) and (10) results in the expansion of the technology from the shaded area in Figure 1 to the area below the broken line *UAW*, and the latter is the new efficient frontier. Department *A* remains efficient, while departments *B* and *C* are no longer efficient and are projected on the units *E* and *F*, respectively. Note that, because of the assumptions about production trade-offs (8) and (10), both target units *E* and *F* are technologically feasible. Therefore, the output radial efficiency of the units *B* and *C* retains its traditional technological meaning. Namely, for each unit the inverse of its output radial efficiency is the ultimate improvement factor by which both of its outputs can be improved.

Example 2. In this example we illustrate the effect of linked production trade-offs on the production technology. For simplicity we consider the case of VRS with a single input (academic staff) and a single output (undergraduate students). The shaded area in Figure 2 corresponds to the VRS technology induced by two departments *A* and *B*. Both departments are efficient in this technology.

Consider the following variants of linked production trade-offs (14) and (16) adapted to our example:

$$P_5^* = (1),\ Q_5^* = (10), \tag{20}$$

$$P_6^* = (-1),\ Q_6^* = (-20). \tag{21}$$

We use the same logic as in Example 1. Starting from unit *A* and applying trade-off (20) in different proportions, we add the ray *AW* to the technology. Similarly, the application of trade-off (21) to unit *A* induces the line *AK*. Using free disposability of input and output, the VRS technology expands to the nonnegative area below the broken line *KAW*, and the latter is its new efficient frontier.

Note that unit *B* is no longer efficient in the expanded technology. Its output radial efficiency is assessed by its projection on the unit *E*. Because the latter unit is producible according to the stated trade-off assumptions, it is a technologically feasible efficient target for department *B*.

5 CRS and VRS Technology with Production Trade-offs

Above we defined production trade-offs as the dual forms of weight restrictions. Their use in the example involving university departments resulted in a meaningful expansion of the CRS and VRS technology and led to a significant improvement of efficiency discrimination.

The missing link in the above development is the definition of technology with production trade-offs. Below we address this gap using the axiomatic approach to the definition of technology pioneered by Banker et al. (1984). The main definitions and results of this section are based on the results of Podinovski (2004a).

5.1 Axiomatic Definitions

The first three axioms are the standard production assumptions that define the conventional VRS technology $\mathcal{T}_{VRS}$. Adding the fourth axiom defines the CRS technology $\mathcal{T}_{CRS}$.

Axiom 1 (Feasibility of observed data). $(X_j, Y_j) \in \mathcal{T}$, for any $j \in J$.

Axiom 2 (Convexity). Technology $\mathcal{T}$ is a convex set.

Axiom 3 (Free disposability). If $(X, Y) \in \mathcal{T}$, $Y \geq Y' \geq 0$ and $X \leq X'$, then $(X', Y') \in \mathcal{T}$.

Axiom 4 (Proportionality). If $(X, Y) \in \mathcal{T}$ and $\alpha \geq 0$, then $(\alpha X, \alpha Y) \in \mathcal{T}$.

The following axiom states that each of the production trade-offs (P_t, Q_t) in (6) can be applied to any unit in technology $\mathcal{T}$, and any number of times (in any proportion) $\pi_t \geq 0$ as long as the resulting unit has nonnegative inputs and outputs.

Axiom 5 (Feasibility of production trade-offs). Let $(X, Y) \in \mathcal{T}$. Then, for each trade-off (P_t, Q_t) in (6) and for any $\pi_t \geq 0$, the unit

$$(\tilde{X}, \tilde{Y}) = (X + \pi_t P_t, Y + \pi_t Q_t) \in \mathcal{T},$$

provided $\tilde{X} \geq 0$ and $\tilde{Y} \geq 0$.

The next, and last, axiom states that the production technology should be a closed set. This is a standard property of production technologies (Shephard 1974, Färe et al. 1985) that is often automatically satisfied and needs not to be stated – this is true in the cases of CRS, VRS and free disposal hull technology of Deprins et al. (1984). However, as shown by an example in Podinovski (2004a), this is not so for technologies that incorporate production trade-offs as stated in Axiom 5. Therefore, the following axiom needs to be explicitly stated.

Axiom 6 (Closedness). Technology $\mathcal{T}$ is a closed set.

The following definition is based on *the minimum extrapolation principle* introduced to DEA by Banker et al. (1984).

Definition 1. The CRS technology $\mathcal{T}_{CRS-TO}$ with trade-offs (6) is the intersection of all technologies $\mathcal{T}$ that satisfy Axioms 1 – 6.

It is straightforward to verify that technology $\mathcal{T}_{CRS-TO}$ satisfies all Axioms 1 – 6. For example, Axiom 2 is satisfied because the intersection of convex sets is a convex set. Definition 1 implies that $\mathcal{T}_{CRS-TO}$ is the smallest technology that satisfies all Axioms 1 – 6. This means that it contains only those DMUs that are required to satisfy the axioms and no other arbitrary units.

The above definition is not constructive, and its equivalent operational statement is given by the following theorem.

Theorem 1 (Podinovski 2004a). Technology $\mathcal{T}_{CRS-TO}$ is the set of all units $(X,Y) \in \mathbb{R}^m_+ \times \mathbb{R}^s_+$ that can be stated in the form

$$Y = \sum_{j=1}^{n} \lambda_j Y_j + \sum_{t=1}^{K} \pi_t Q_t - e, \tag{22}$$

$$X = \sum_{j=1}^{n} \lambda_j X_j + \sum_{t=1}^{K} \pi_t P_t + d, \tag{23}$$

where $\lambda = (\lambda_1,...,\lambda_n) \in \mathbb{R}^n_+$, $\pi = (\pi_1,...,\pi_K) \in \mathbb{R}^K_+$, $e \in \mathbb{R}^s_+$ and $d \in \mathbb{R}^m_+$.

Theorem 1 provides a meaningful interpretation to the envelopment models (4) and (5). It shows that the radial improvement of the input and, respectively, output vectors of the unit (X_o, Y_o) is performed within the technology $\mathcal{T}_{CRS-TO}$. Note, however, that this interpretation is correct only if the improved unit has nonnegative input and output vectors, as required by Theorem 1. This requirement is automatically satisfied in model (5) because the output-improvement factor η is maximised. In model (4) the input-improvement factor θ is minimised and may in some cases become negative. It may appear that we need to add the condition $\theta \geq 0$ to the constraints of model (4) – this would remedy the problem and guarantee that the minimisation of θ is performed in technology $\mathcal{T}_{CRS-TO}$. While this is possible, there are two reasons why this may not be a good idea.

First, adding the condition $\theta \geq 0$ to the constraints of model (4) would invalidate its duality with the multiplier model (2). The second and, perhaps, more important consideration is that the feasibility of negative values of θ in model (4) indicates an inconsistency within the trade-offs (6) or, equivalently, weight restrictions (1). Allowing θ to take on negative values in the envelopment models make them self-testing for errors in the construction of trade-offs (or weight restrictions). We consider this issue in detail in the next section.

Generally though, the nonnegativity conditions $(X,Y) \in \mathbb{R}^m_+ \times \mathbb{R}^s_+$ are important in the statement of technology $\mathcal{T}_{CRS-TO}$ and should not be omitted unless proved redundant in a particular DEA model. This is discussed further in Section 7 (see Remark 2) in relation to the additive DEA model based on the above technology.

In the case of VRS, we follow the same logic as above and give the following definition.

Definition 2. The VRS technology $\mathcal{T}_{VRS-TO}$ with trade-offs (6) is the intersection of all technologies $\mathcal{T}$ that satisfy Axioms 1 – 3, 5 and 6.

As in the above case, it is straightforward to verify that technology $\mathcal{T}_{VRS-TO}$ satisfies Axioms 1 – 3, 5 and 6 and is, therefore, the smallest technology that satisfies them.

Theorem 2 (Podinovski 2004a). Technology $\mathcal{T}_{VRS-TO}$ is the set of all units $(X,Y) \in \mathbb{R}_+^m \times \mathbb{R}_+^s$ that can be stated in the form (22) and (23), subject to the additional normalising equality (7) and the same nonnegativity conditions on vectors λ, π, e and d as in Theorem 1.

The duality of weight restrictions and production trade-offs allows us to give a positive answer to the long-standing question of whether the use of weight restrictions in VRS DEA models is theoretically sound (Thanassoulis and Allen 1998). The counter-argument is that in CRS models the marginal rates of transformation and substitution between inputs and outputs (that define the slopes of facets on the boundary of the technology) are invariant with respect to the scaling (or the size) of the unit, while in the VRS technology this is not so. The main concern is then that the weight restrictions that specify bounds on the marginal rates would be inappropriate in the VRS technology because such rates change with the scale of operations. This argument is weakened by the fact that the marginal rates in the CRS technology are still generally different at any two units, unless one is a scaled variant of the other.

The above problem does not arise if we interpret weight restrictions as the dual forms of production trade-offs and assess the latter in the first place. Indeed, if production trade-offs (6) are assumed technologically feasible in the CRS technology $\mathcal{T}_{CRS-TO}$ (in the sense of Axiom 5), then they must be technologically feasible in the VRS technology $\mathcal{T}_{VRS-TO}$ because the latter is a subset of $\mathcal{T}_{CRS-TO}$. Therefore, any production trade-offs (or weight restrictions based on them) that are deemed realistic and appropriate in the CRS model, are also acceptable and can be used in the VRS model.

5.2 Some Properties of CRS and VRS Technologies with Trade-offs

Below we establish two properties of technologies $\mathcal{T}_{CRS-TO}$ and $\mathcal{T}_{VRS-TO}$.

Theorem 3. Technologies $\mathcal{T}_{CRS-TO}$ and $\mathcal{T}_{VRS-TO}$ are polyhedral sets. In particular, $\mathcal{T}_{CRS-TO}$ is a polyhedral cone.

Proof of Theorem 3. The set P of all solutions $\{X,Y,\lambda,\pi,e,d\}$ to the set of linear equations (22), (23) and inequalities $X,Y,\lambda,\pi,e,d \geq 0$ is a polyhedral set in $\mathbb{R}^{2(m+s)+n+K}$. Technology $\mathcal{T}_{CRS-TO}$ in Definition 1 is the projection of P on its input and output dimensions X and Y.

By the known projection lemma (see, e.g., Jones et al. 2008, Lemma 3.1), $\mathcal{T}_{CRS-TO}$ is a polyhedral set. Because $\mathcal{T}_{CRS-TO}$ satisfies Axiom 4, it is a cone. The case of technology $\mathcal{T}_{VRS-TO}$ is considered in a similar way. ■

The second property is somewhat more subtle. Without production trade-offs, the conventional CRS technology is the cone extension of the VRS technology. This means that any unit (X,Y) in the CRS technology is obtained by the scaling of some unit $(\tilde{X},\tilde{Y})$ in the VRS technology by some factor $\alpha \geq 0$. This result is generally incorrect for the CRS and VRS technologies that incorporate production trade-offs (although it is “almost correct” in the sense defined below).

Example 3. Consider the CRS and VRS technologies with a single input and single output induced by the single observed unit $A=(2,1)$. Suppose we specified the linked trade-off: $(P,Q)=(1,2)$. Figure 3 shows the resulting VRS technology $\mathcal{T}_{VRS-TO}$ as the shaded area below the broken line *GAF*. Note that the ray *AF* is obtained by the application of trade-off (P,Q) to the unit *A* following the same logic as in Example 2. Furthermore, the CRS technology $\mathcal{T}_{CRS-TO}$ is the cone under the ray *OE*: the ray *OE* is obtained by the application of trade-off (P,Q) to the zero unit – the latter is included in the original CRS technology. This implies that, for example, unit $B=(1,2)$ is in technology $\mathcal{T}_{CRS-TO}$. (As an alternative argument, unit *B* satisfies the conditions of Theorem 1 with $\lambda_1=0$ and $\pi_1=1$.) It is, however, straightforward to show that there exists no unit $(\tilde{X},\tilde{Y}) \in \mathcal{T}_{VRS-TO}$ and $\alpha \geq 0$ such that $B=\alpha(\tilde{X},\tilde{Y})$.

Example 3 shows that technology $\mathcal{T}_{CRS-TO}$ is generally not the cone extension of $\mathcal{T}_{VRS-TO}$. Below we prove that $\mathcal{T}_{CRS-TO}$ is the *closed* cone extension of $\mathcal{T}_{VRS-TO}$. To state this formally, denote the cone extension of $\mathcal{T}_{VRS-TO}$ as

$$cone\, \mathcal{T}_{VRS-TO} = \left\{ (X,Y) \in \mathbb{R}^m \times \mathbb{R}^s \mid \exists (\tilde{X},\tilde{Y}) \in \mathcal{T}_{VRS-TO}, \alpha \geq 0 : (X,Y) = \alpha(\tilde{X},\tilde{Y}) \right\}.$$

Denote $cl(cone\, \mathcal{T}_{VRS-TO})$ the closure of the set $cone\, \mathcal{T}_{VRS-TO}$ (intersection of all closed sets containing $cone\, \mathcal{T}_{VRS-TO}$).

Theorem 4. Technology $\mathcal{T}_{CRS-TO}$ is the closed cone induced by $\mathcal{T}_{VRS-TO}$:

$$\mathcal{T}_{CRS-TO} = cl(cone\, \mathcal{T}_{VRS-TO}).$$

Proof of Theorem 4. By Theorem 2, any $(\tilde{X},\tilde{Y}) \in \mathcal{T}_{VRS-TO}$ satisfies (22), (23) and (7) with some vectors $\tilde{\lambda}$, $\tilde{\pi}$, $\tilde{e}$ and $\tilde{d}$. For any $\alpha \geq 0$, $\alpha(\tilde{X},\tilde{Y})$ satisfies (22) and (23) with the vectors $\alpha\tilde{\lambda}$, $\alpha\tilde{\pi}$, $\alpha\tilde{e}$ and $\alpha\tilde{d}$. By Theorem 1, $(\tilde{X},\tilde{Y}) \in \mathcal{T}_{CRS-TO}$. Therefore, $cone\, \mathcal{T}_{VRS-TO} \subseteq \mathcal{T}_{CRS-TO}$, and $cl(cone\, \mathcal{T}_{VRS-TO}) \subseteq cl\, \mathcal{T}_{CRS-TO} = \mathcal{T}_{CRS-TO}$. (The last equality is true because $\mathcal{T}_{CRS-TO}$ satisfies Axiom 6.)

Conversely, let $(X,Y) \in \mathcal{T}_{CRS-TO}$. Then (X,Y) satisfies (22) and (23) with some vectors λ', π', e' and d'. Let $\lambda^* = \sum_{j=1}^{n} \lambda'_j$. Two cases arise.

Case 1. Assume that $\lambda^* > 0$. Define $(\tilde{X},\tilde{Y}) = (1/\lambda^*)(X,Y)$. Then $(\tilde{X},\tilde{Y}) \in \mathcal{T}_{VRS-TO}$ because it satisfies (22), (23) and (7) with $\lambda = \lambda'/\lambda^*$, $\pi = \pi'/\lambda^*$, $e = e'/\lambda^*$ and $d = d'/\lambda^*$. Because $(X,Y) = \alpha(\tilde{X},\tilde{Y})$ where $\alpha = \lambda^*$, we have $(X,Y) \in cone\, \mathcal{T}_{VRS-TO} \subseteq cl(cone\, \mathcal{T}_{VRS-TO})$.

Case 2. Assume that $\lambda^* = 0$. Therefore, $\lambda' = 0$. (This is the case for unit *B* in Example 3.) Consider the sequence of units (X_k,Y_k), $k = 1,2,\ldots$, defined as follows:

$$(X_k,Y_k) = \sum_{j=1}^{n} \left(\frac{1}{n}(X_j,Y_j) \right) + k(X,Y). \tag{24}$$

Because both terms on the right-hand side of (24) are nonnegative, each unit (X_k,Y_k) is nonnegative. It is straightforward to verify that (X_k,Y_k) satisfies conditions (22), (23) and (7) with the vector λ_k whose components are $(\lambda_k)_j = 1/n$, $j = 1,\ldots,n$, and vectors $\pi_k = k\pi'$, $e_k = ke'$ and $d_k = kd'$. Therefore, $(X_k,Y_k) \in \mathcal{T}_{VRS-TO}$, for all $k = 1,2,\ldots$

Define the sequence of units $\left(\tilde{X}_k,\tilde{Y}_k\right) = (1/k)\left(X_k,Y_k\right)$. Obviously, we have $\left(\tilde{X}_k,\tilde{Y}_k\right) \in cone\, \mathcal{T}_{VRS-TO}$, for all *k*. Note that (X,Y) is the limit unit of the sequence of units $\left(\tilde{X}_k,\tilde{Y}_k\right)$. Indeed, based on (24),

$$(\tilde{X}_k,\tilde{Y}_k) = \frac{1}{k}\sum_{j=1}^{n} \left(\frac{1}{n}(X_j,Y_j) \right) + (X,Y) \underset{k\to+\infty}{\longrightarrow} (X,Y).$$

Therefore $(X,Y) \in cl(cone\, \mathcal{T}_{VRS-TO})$. Because (X,Y) is an arbitrary unit in $\mathcal{T}_{CRS-TO}$, in both cases 1 and 2 we have $\mathcal{T}_{CRS-TO} \subseteq cl(cone\, \mathcal{T}_{VRS-TO})$. Taking into account the inverse embedding obtained in the first part of the proof, we have $\mathcal{T}_{CRS-TO} = cl(cone\, \mathcal{T}_{VRS-TO})$. ■

Theorem 4 states that the CRS technology $\mathcal{T}_{CRS-TO}$ is obtained from the VRS technology $\mathcal{T}_{VRS-TO}$ by the scaling of its units by all factors $\alpha \geq 0$, and subsequently adding all limit points (units) to the resulting set.

6 Weight Restrictions and the Infeasibility Problem

It is well-known that the use of weight restrictions in multiplier models (2) and (3), and in their VRS analogues, may result in their infeasibility (see, e.g., Allen et al. 1997, Pedraja-Chaparro et al. 1997). A similar problem may occur when production trade-offs are incorporated in envelopment DEA models. By duality, if a multiplier model with weight restrictions is infeasible, its dual envelopment model (which is always feasible) must have an unbounded objective function.

The unboundness of the objective function η in the output-maximisation CRS model (5) and its VRS analogue indicates that the incorporation of weight restrictions (production trade-offs) has created an unlimited production of the output vector Y_o. (Because ηY_o can be taken to infinity while keeping the input vector X_o constant.) This is inconsistent with the established properties of production technologies (Shephard 1974, Färe et al. 1985) and indicates that an error has occurred in the construction of weight restrictions or trade-offs.

The unboundness of the objective function θ in the input-minimisation model (4) or its VRS analogue implies that $\theta = 0$ is feasible in the model. Consequently, the technology allows free production of the output vector Y_o from the zero vector of inputs $\theta X_o = 0 X_o$. This is an equally problematic situation that indicates that weight restrictions should be reconsidered.

In the author's experience based on teaching DEA to a large class of undergraduate students for many years, who were asked to use weight restrictions in their work, the above infeasibility problems are not unusual. These are more likely to happen if the model incorporates a relatively large number of weight restrictions of complex structure: those that involve several input and output weights in one inequality, as in (15) and (17). The use of trade-offs for the assessment of weight restrictions facilitates and often encourages the formulation of complex weight restrictions. For example, weight restrictions (15) and (17) that have a clear meaning as stated in Assumptions 5 and 6 are unlikely to be stated using value judgements, because it may not even be clear what they mean in value terms.

Podinovski and Bouzdine-Chameeva (2013) show that free and unlimited production of output vectors may occur *even if all multiplier models are feasible and all efficiency scores appear plausible*. In such cases, the technology is modelled incorrectly and the efficiency scores are also incorrect. One cannot therefore rely on the fact that the efficiency scores appear unproblematic – there may still be an undetected underlying problem with weight restrictions that invalidates the results of analysis and needs correcting.

Below we outline the results presented in Podinovski and Bouzdine-Chameeva (2013). These include a description of the infeasibility (and unboundness) problem caused by weight restrictions and the forms it can take, depending on the assumption of returns to scale (VRS or CRS) and the orientation of the model (input minimisation or output maximisation). This leads to the formulation of analytical and computational tests that give us a conclusive answer as to whether there is a problem with weight restrictions.

6.1 Definitions and Examples

We start with the following two definitions. Let $Y_o \in \mathbb{R}^s_+$, $Y_o \neq 0$, be a vector of outputs.

Definition 3. Technology $\mathcal{T}$ allows *free production* of vector Y_o if $(0, Y_o) \in \mathcal{T}$.

Definition 4. Technology $\mathcal{T}$ allows *unlimited production* of vector Y_o if there exists a vector of inputs X_o such that $(X_o, \alpha Y_o) \in \mathcal{T}$ for all $\alpha \geq 0$.

Podinovski and Bouzdine-Chameeva (2013) prove that the above two notions are equivalent in any cone technology, e.g., in technology $\mathcal{T}_{CRS-TO}$: the existence of free production implies the existence of unlimited production, and vice versa. In a non-cone technology, e.g., in $\mathcal{T}_{VRS-TO}$, the two notions are generally different. Furthermore, in any convex technology (e.g., in $\mathcal{T}_{CRS-TO}$ and $\mathcal{T}_{VRS-TO}$), the specification of vector X_o in Definition 4 is unimportant: if vector Y_o can be produced in an unlimited quantity α from the input vector X_o, then it can be produced in an unlimited quantity from the input vector X of any other unit (X, Y) in the technology.

It is straightforward to verify that, under the nonnegativity assumptions made about the observed DMUs, conventional CRS and VRS production technologies do not allow free or unlimited production of output vectors, but the incorporation of weight restrictions (production trade-offs) may create it. The following two examples demonstrate this effect.

Example 4. Suppose we made a mistake in the assessment of production trade-offs (8) and (10), and stated them as follows:

$$\tilde{P}_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix},\ \tilde{Q}_1 = \begin{pmatrix} 4 \\ -1 \\ 0 \end{pmatrix}, \tag{25}$$

$$\tilde{P}_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix},\ \tilde{Q}_2 = \begin{pmatrix} -3 \\ 2 \\ 0 \end{pmatrix}. \tag{26}$$

It is easy to see that the above trade-offs induce unlimited production of the two outputs (undergraduate and master students) in the VRS and CRS technology. Figure 4 is a modification of Figure 1 to this case.

Starting from unit *A* and applying trade-off (25) 100 times, we substitute 100 master students by 400 undergraduate students. This creates point E_1 on the graph. Subsequently applying trade-off (26) 100 times, we substitute 300 undergraduate students by 200 master students. The resulting unit A_1 has 100 more of both types of student compared to the original unit *A*, and "achieves" this without any extra input. We can continue this process and generate a further sequence of units A_2, A_3,…, taking the production of outputs to infinity. (The lightly shaded area in Figure 4 shows the region of units dominated by A_3. By free disposability of output, this region is also included in the technology. As the sequence of units A_t, $t = 1, 2, ...$, tends to infinity, the corresponding dominated area covers the whole nonnegative orthant.)

Example 5. Consider the VRS technology as in Figure 2. Assume we replaced the production trade-off (21) by the following trade-off:

$$\tilde{P} = (-1),\ \tilde{Q} = (-10). \tag{27}$$

Figure 5 shows the effect of trade-off (27) on the VRS technology. Starting at unit *A* and consecutively applying this trade-off, we generate the line *AK* which, together with the region below it, should be added to the technology. Note that unit *K* has a zero input and a strictly positive output. This means that the expanded technology allows free production and indicates that trade-off (27) should be reconsidered.

Note that the above problem cannot be observed by the efficiency calculations: the output radial efficiency of departments *A* and *B* in this example is equal to 1 and 0.5, respectively, and is not suspicious. However, because the slope of the efficient boundary *KA* is incorrect, the calculated efficiencies are also incorrect.

6.2 Theoretical Results

Below we give a complete characterisation of problematic outcomes in the CRS and VRS DEA models with weight restrictions (production trade-offs) that are caused by free or unlimited production of vector Y_o in the corresponding technology. If any of such outcomes are observed in practical computations, this implies that an error has occurred in the assessment of weight restrictions (or, equivalently, production trade-offs), and these need to be reconsidered.

The first theorem deals with the case of CRS.

Theorem 5 (Podinovski and Bouzdine-Chameeva 2013). Let $(X_o, Y_o) \in \mathcal{T}_{CRS-TO}$ and let $X_o \neq 0$ and $Y_o \neq 0$. (For example, (X_o, Y_o) may be an observed unit.) Then the following three statements are equivalent:

(a) There exists free and unlimited production of output vector Y_o in technology $\mathcal{T}_{CRS-TO}$.

(b) The CRS input-minimisation envelopment model $\mathbb{E}^1_{CRS}$ is unbounded or has a finite optimal value $\theta^* = 0$. Its dual multiplier model $\mathbb{M}^1_{CRS}$ is infeasible or has an optimal value $\theta^* = 0$, respectively.

(c) The CRS output-maximisation envelopment model $\mathbb{E}^2_{CRS}$ is unbounded. Its dual multiplier model $\mathbb{M}^2_{CRS}$ is infeasible.

The next result deals with the case of VRS. Because in this technology the notions of free and unlimited production are generally not equivalent, these are considered separately.

Theorem 6 (Podinovski and Bouzdine-Chameeva 2013). Let $(X_o, Y_o) \in \mathcal{T}_{VRS-TO}$ and let $X_o \neq 0$ and $Y_o \neq 0$. (For example, (X_o, Y_o) may be an observed unit.) Then the following statements are true:

(a) There exists free production of output vector Y_o in technology $\mathcal{T}_{VRS-TO}$ if and only if the VRS input-minimisation envelopment model $\mathbb{E}^1_{VRS}$ is either unbounded or has a finite optimal value $\theta^* \leq 0$. Its dual multiplier model $\mathbb{M}^1_{VRS}$ is, respectively, infeasible or has a finite optimal value $\theta^* \leq 0$.

(b) There exists unlimited production of output vector Y_o in technology $\mathcal{T}_{VRS-TO}$ if and only if the VRS output-maximisation multiplier model $\mathbb{E}^2_{VRS}$ is unbounded. Its dual multiplier model $\mathbb{M}^2_{VRS}$ is infeasible.

One of the differences between the cases of CRS and VRS highlighted by Theorems 5 and 6 is that free production in the VRS technology may result in a finite negative value of the input efficiency θ^*. For example, consider unit $G = (50, 50)$ in the VRS technology in Figure 5. The input radial projection of G is $H = (-5, 50)$. Solving the envelopment model $\mathbb{E}^1_{VRS}$ produces the finite value $\theta^* = -5/50 = -0.1$ and illustrates part (a) of Theorem 6.

The above two theorems do not solve the problem of identifying problematic weight restrictions (trade-offs) completely: even if no problematic outcomes occur with the assessment of all observed units (X_j, Y_j), this guarantees only that there is no free or unlimited production of the output vectors Y_j of observed units. This does not however guarantee that there is no free or unlimited production of other output vectors in the technology. For example, in the case of VRS technology in Figure 5, Theorem 6 would not identify any problem when the input or output radial efficiency of both observed units *A* and *B* is assessed.

Podinovski and Bouzdine-Chameeva (2013) suggest two approaches, analytical and computational, that allow us to examine if the incorporation of production trade-offs (weight restrictions) has induced free or unlimited production in the technology. This task is simplified by the following statement.

Theorem 7 (Podinovski and Bouzdine-Chameeva 2013). The existence of free (and therefore unlimited production) of the output vector Y_o in technology $\mathcal{T}_{CRS-TO}$ is equivalent to the existence of either free or unlimited production of vector Y_o (but not necessarily both) in technology $\mathcal{T}_{VRS-TO}$.

According to Theorem 7, if there is a problem with free or unlimited production in either CRS or VRS technology, then there is a similar problem in the other. Because the notions of free and unlimited production are equivalent in the CRS technology, and also because the choice of vector *X* is unimportant for the latter notion, it suffices to test for the existence of unlimited production with the input vector *X* of an arbitrary unit (X, Y) in the CRS technology $\mathcal{T}_{CRS-TO}$.

Podinovski and Bouzdine-Chameeva (2013) consider two cases. The simpler case arises if weight restrictions (1) are not linked. In this case the testing is reduced to verifying a simple algebraic condition. If weight restrictions (1) include linked restrictions, the testing is performed by solving specially constructed linear programs. Below we outline the two cases.

6.3 Free Production with Not Linked Trade-offs

The most straightforward case arises if the weight restrictions are not linked. Then (1) can be restated as follows:

$$u^{\top} Q_t \leq 0, \qquad t = 1, ..., K_1, \tag{28}$$

$$-v^{\top} P_t \leq 0, \qquad t = 1, ..., K_2. \tag{29}$$

Theorem 8 (Podinovski and Bouzdine-Chameeva 2013). Technology $\mathcal{T}_{CRS-TO}$ does not allow free (and unlimited) production if and only if both of the following two conditions are satisfied:

(a) there exists a strictly positive vector $u^* > 0$ that satisfies (28);

(b) there exists a nonnegative vector $v^* \geq 0$ that satisfies (29) such that $(v^*)^{\top} X_j > 0$ holds for all observed units j = 1, ..., n.

(If either group of weight restrictions (28) or (29) is missing, then the corresponding condition (a) or (b) is removed from the above statement.)

Note that the vectors u^* and v^* do not need to satisfy the conditions of models (2) or (3) – all that is required is that such vectors satisfy (28) and (29).

In practical applications all inputs of all observed DMUs $j = 1, ..., n$ are usually strictly positive. In this case condition (b) of Theorem 8 is equivalent to the simpler condition: there exists a nonzero vector $v^* \geq 0$ that satisfies (29).

If some of the inputs of observed DMUs are equal to zero, the above simplified condition does not apply. However, to prove that there is no free production, a simpler *sufficient condition* may be used. (Obviously, if it is not satisfied, this does not mean that there is free production – we need to use Theorem 8 for a definitive answer.)

Corollary 1. If there exist strictly positive vectors $u^* > 0$ and $v^* > 0$ that satisfy (28) and (29), then technology $\mathcal{T}_{CRS-TO}$ does not allow free (and unlimited) production.

As an illustration, refer to Example 1 in which we used the trade-offs between undergraduate and master students as stated in (8) and (10). The resulting technology was illustrated in Figure 1. The corresponding weight restrictions (9) and (11) are simultaneously satisfied, for example, by strictly positive weights $u_1 = u_2 = 1$. This means that condition (a) of Theorem 8 is true. Because there are no weight restrictions involving input weights, condition (b) of Theorem 8 should be ignored. By Theorem 8 or its Corollary 1, the two trade-offs (8) and (10) do not cause free or unlimited production in either CRS or VRS technology, which is consistent with Figure 1.

Let us illustrate how Theorem 8 can be used to detect free production when it exists, even if all efficiency scores appear unproblematic.

Example 6. In Example 4 we showed how the use of trade-offs (25) and (26) resulted in the unlimited production of two outputs (undergraduate and master students). If we use the same two trade-offs with the data set in Table 1, they induce unlimited production of the two outputs in the same way but the problem *is not observed* from the efficiency calculations and becomes hidden.

Table 6 shows the efficiency scores (in %) in the CRS and VRS DEA models for the departments as in Table 1. Both the CRS and VRS models incorporate only two production trade-offs (25) and (26). (These models are obtained from the models CRS 2 and VRS 2 discussed above in which the "good" trade-offs (8) and (10) are replaced by the problematic trade-offs (25) and (26).)

Note that the results of computations in Table 6 do not appear problematic – the only exception may be the unusually low "efficiency" of department D2 in both models. In such cases it is easy to miss the underlying problem. To see if there is a problem we use Theorem 8 and restate production trade-offs (25) and (26) as the weight restrictions

$$4u_1 - u_2 \le 0, \tag{30}$$

$$-3u_1 + 2u_2 \le 0. \tag{31}$$

It is straightforward to show that the above inequalities cannot be satisfied by strictly positive weights u_1 and u_2. Indeed, adding the two inequalities (30) and (31), we obtain $u_1 + u_2 \le 0$, which does not allow a strictly positive solution vector. By Theorem 8, production trade-offs (25) and (26) induce free (and unlimited) production in the CRS technology, and the CRS efficiency scores are, although plausible, obviously meaningless. By Theorem 7, the efficiency scores in the VRS model are also incorrect.

6.4 Free Production with Linked Trade-offs

In the general case of linked weight restrictions (1) Podinovski and Bouzdine-Chameeva (2013) develop two computational procedures to test if there is free (and unlimited) production in the CRS technology. Below we describe one of them.

The idea of this method is simple and based on the following fact: technology $\mathcal{T}_{CRS-TO}$ allows an unlimited production of a vector Y_o *if and only if* it allows an unlimited production of each of its individual *positive* outputs, provided all the other individual outputs are taken equal to zero. (The "only if" part of this statement is obvious. The "if" part follows from the following. Suppose the technology allows the production of each individual output $(Y_o)_r$, $r=1,...,s$, in any proportion $\alpha \geq 0$, from the input vector X_o. Then the simple average of all s such units, each producing the single output $\alpha(Y_o)_r$, is the unit $(X_o,(\alpha/s)Y_o) \in \mathcal{T}_{CRS-TO}$. Because s is constant and α is arbitrarily large, technology $\mathcal{T}_{CRS-TO}$ allows an unlimited production of vector Y_o.)

The above suggests that we can test for unlimited production as follows. First, we select any (e.g., observed) unit $(X_o,Y_o) \in \mathcal{T}_{CRS-TO}$ such that all components of vector Y_o are strictly positive: $(Y_o)_r > 0$, for all $r=1,...,s$. If no such observed unit exists, we can always take the simple average of all observed units. Because each output *r* is strictly positive for at least one observed unit *j*, the average of all observed units will have a strictly positive output vector.

Define s artificial output vectors U_r, $r=1,...,s$, as follows. Each of these vectors has only one positive component:

$$U_1 = \left((Y_o)_1, 0, ..., 0\right)^\top, \ \ldots, \ U_s = \left(0, ..., 0, (Y_o)_s\right)^\top .$$

Consider s DMUs in the form (X_o,U_r), where $r=1,...,s$. Each of such units is dominated by the original unit (X_o,Y_o) and therefore $(X_o,U_r) \in \mathcal{T}_{CRS-TO}$. We can now expand the set of observed DMUs J by incorporating the above s artificial units. Because the latter units are dominated, the technology $\mathcal{T}_{CRS-TO}$ remains unchanged.

We now solve *s* output-maximisation multiplier models, one for each unit (X_o,U_ρ), $\rho=1,...,s$. (We use index ρ to differentiate from r in the same formulation.)

$$\eta^* = \min \quad v^\top X_o, \tag{32}$$

$$\begin{aligned} \text{subject to} \quad & u^\top U_\rho = 1, \\ & u^\top Y_j - v^\top X_j \leq 0, \qquad j=1,...,n, \\ & u^\top U_r - v^\top X_o \leq 0, \qquad r=1,...,s, \end{aligned}$$

$$u^{\top}Q_t - v^{\top}P_t \le 0, \qquad t = 1,...,K,$$

$$u, v \ge 0.$$

Theorem 9 (Podinovski and Bouzdine-Chameeva 2013). Technology $\mathcal{T}_{CRS-TO}$ allows free (and unlimited) production if and only if there exists a $\rho = 1,...,s$ such that the multiplier model (32) is infeasible.

Obviously, instead of model (32), we can solve its dual envelopment model. In this case the infeasibility of model (32) is equivalent to the unboundness of the envelopment model. Also note that the constraints $u^{\top}U_r - v^{\top}X_o \le 0$ in model (32) are redundant and can in principle be removed because, as discussed, units (X_o, U_r) are dominated. From the practical point of view, however, it may be beneficial to keep model (32) as stated, because in this case it can be solved by standard DEA solvers.

Example 7. As an illustration, consider the university departments in Table 1 and the eight trade-offs discussed in Section 3. Because some of these trade-offs are linked, we use the method based on program (32) to verify that the combination *of this particular data set* and trade-offs does not induce free or unlimited production.

As the starting point, let us choose department D1 as the unit (X_o, Y_o). (Alternatively, we can choose any department from D1 to D6 for this purpose, but not D7 because its second output is zero.) Following the above procedure, define three artificial units with the vector of inputs $X_o = (92,15)^{\top}$ as in department D1, and the following different output vectors:

$$U_1 = (800,0,0)^{\top},\ U_2 = (0,200,0)^{\top},\ U_3 = (0,0,90)^{\top}.$$

We now add the three units (X_o, U_1), (X_o, U_2) and (X_o, U_3) to the set of departments D1 – D7. Because all three additional departments are dominated by D1, the technology does not change. Finally, we assess the output radial efficiency of the three additional departments in the CRS multiplier model (32). The corresponding three optimal values of program (32) are finite and equal, respectively, to 3.632, 6.041 and 2.037. (The output radial efficiency of the three artificial units is, respectively, 0.2753, 0.1655 and 0.4909. The output radial efficiency of departments D1 – D7, if calculated simultaneously by the software, is the same as without the additional three units.) By Theorem 9, the CRS (and consequently VRS) technology based on the data set in Table 1 and the eight trade-offs does not allow free or unlimited production.

7 Solving DEA Models with Production Trade-offs

Conventional CRS and VRS DEA models (without weight restrictions) are usually solved using either a two-stage computational procedure or an analogous single-stage method utilizing a non-Archimedean ε (in practice taken equal to a very small positive number). These methods are summarized in Thanassoulis et al. (2008) and Cooper et al. (2011b).

In many applications of DEA only the radial efficiency of the DMUs is of interest, and the first stage of the two-stage method suffices for this purpose. It identifies the radial projection of the assessed DMU on the boundary of the VRS or CRS technology and produces the DMU's radial input or output efficiency. Because the radial projection of an inefficient DMU may be only *weakly* efficient, the identification of its efficient target (in the Pareto sense) requires the second optimisation stage in which the sum of input and output slacks is maximised. Performing the second stage identifies the efficient target of the DMU and the reference set of its efficient peers. The latter are the observed DMUs j that have a corresponding multiplier $\lambda_j > 0$ in the optimal solution to the second-stage linear program.

Podinovski (2007b) shows that the application of the standard second stage to DEA models with weight restrictions (or production trade-offs) may result in a target unit with meaningless negative values of some inputs. (This is unrelated to the issue of inconsistent weight restrictions discussed in the previous section.) In the suggested corrected procedure, the conventional second stage is split into two new stages, and the complete solution method becomes a three-stage procedure. Depending on the purpose of a DEA study, only the first, two first or all three computational stages may need to be performed.

Below we outline these three stages. We assume that the weight restrictions (production trade-offs) have already been checked using the methods described in the previous section, and that the underlying VRS or CRS technology does not allow free or unlimited production of non-zero output vectors.

Stage 1 (Assessing the radial efficiency). This task is straightforward and requires the solution of the appropriate CRS or VRS envelopment model, or their dual multiplier forms, as stated in Section 2.

Stage 2 (Identifying efficient targets). An efficient target of DMU (X_o, Y_o) is obtained by solving the specially constructed additive DEA model formulated in Section 7.2 below.

Stage 3 (Identifying reference sets of efficient peer units). This stage is required because, even if the multiplier λ_j is strictly positive in an optimal solution to the model used at

Stage 2, the corresponding observed DMU *j* may be inefficient. An example of this is given in Podinovski (2007b). The linear program solved at Stage 3 is presented in Section 7.3.

7.1 Stage 1: Assessing the radial efficiency

Most applications of DEA are concerned only with the input or output radial efficiency of the units. In such applications this stage is the only one that needs performing. Depending on the assumption of CRS or VRS and the orientation of the model (input minimisation or output maximisation), the radial efficiency of DMU (X_o, Y_o) is assessed by solving the corresponding envelopment (or multiplier) model stated in Section 2.

This stage also identifies the radial projection (target) unit (X^*, Y^*) of the DMU (X_o, Y_o). In the case of input minimisation, $(X^*, Y^*) = (\theta^* X_o, Y_o)$, where θ^* is the input radial efficiency of DMU (X_o, Y_o). In the case of output maximisation, $(X^*, Y^*) = (X_o, \eta^* Y_o)$, where η^* is the inverse output radial efficiency of DMU (X_o, Y_o). (The value η^* is the optimal value in the corresponding envelopment and multiplier models that is inverse to the output efficiency measure.)

7.2 Stage 2: Identifying Efficient Targets

As in the case of conventional CRS and VRS DEA models, this stage should be performed only if we need to identify efficient targets of individual DMUs. In particular, the computations at this stage do not alter the radial efficiency assessed at Stage 1.

The need of the second stage arises because the radial target (X^*, Y^*) assessed at Stage 1 may be a weakly efficient unit and not efficient in the Pareto sense. The conventional second optimisation stage aims at maximising the sum of input and output slacks that improve the unit (X^*, Y^*). The same idea is applicable to DEA models with weight restrictions (production trade-offs), but an additional care has to be taken of the nonnegativity of inputs in the resulting efficient unit (which is automatically maintained in the standard models without weight restrictions).

The following program identifies possible individual improvements to the inputs and outputs of the unit (X^*, Y^*):

$$\sigma^* = \max \quad \sum_{r=1}^{s} \varepsilon_r + \sum_{i=1}^{m} \delta_i \,, \tag{33}$$

subject to $\quad (X^* - \delta, Y^* + \varepsilon) \in \mathcal{T}$, where $\varepsilon \in \mathbb{R}_+^s$, $\delta \in \mathbb{R}_+^m$, and technology $\mathcal{T}$ is either $\mathcal{T}_{CRS-TO}$ or $\mathcal{T}_{VRS-TO}$.

To be specific, consider the case of CRS. Based on Theorem 1, program (33) takes on the form

$$\sigma^* = \max \quad \sum_{r=1}^{s} \varepsilon_r + \sum_{i=1}^{m} \delta_i , \tag{34.1}$$

$$\text{subject to} \quad \sum_{j=1}^{n} \lambda_j Y_j + \sum_{t=1}^{K} \pi_t Q_t - e = Y^* + \varepsilon , \tag{34.2}$$

$$\sum_{j=1}^{n} \lambda_j X_j + \sum_{t=1}^{K} \pi_t P_t + d = X^* - \delta , \tag{34.3}$$

$$Y^* + \varepsilon \geq 0 , \tag{34.4}$$

$$X^* - \delta \geq 0 , \tag{34.5}$$

$$\lambda, \pi, e, d, \varepsilon, \delta \geq 0 . \tag{34.6}$$

Note that program (34) can be simplified. First, at any of its optimal solutions the vector e must be a zero vector. Indeed, if we assume the converse ($e \geq 0$ and $e \neq 0$) then redefining $\tilde{e} = 0$ and $\tilde{\varepsilon} = \varepsilon + e$ keeps (34.2) true and improves the objective function (34.1), which is impossible due to the assumed optimality of the current solution. Therefore, vector e in program (34) can be assumed zero and removed from the formulation. Second, condition (34.4) is redundant because both vectors Y^* and ε are nonnegative.

The resulting model is as follows:

$$\sigma^* = \max \quad \sum_{r=1}^{s} \varepsilon_r + \sum_{i=1}^{m} \delta_i , \tag{35.1}$$

$$\text{subject to} \quad \sum_{j=1}^{n} \lambda_j Y_j + \sum_{t=1}^{K} \pi_t Q_t = Y^* + \varepsilon , \tag{35.2}$$

$$\sum_{j=1}^{n} \lambda_j X_j + \sum_{t=1}^{K} \pi_t P_t + d = X^* - \delta , \tag{35.3}$$

$$X^* - \delta \geq 0 , \tag{35.4}$$

$$\lambda, \pi, d, \varepsilon, \delta \geq 0 . \tag{35.5}$$

Model (3) is the same as model (6) stated in Podinovski (2007b). In the latter model the above condition (35.4) is replaced by an equivalent requirement that the expression on the left-hand side of equality (35.3) is nonnegative.

As already stated, we assume that technology $\mathcal{T}_{CRS-TO}$ does not allow free and unlimited production. Therefore the objective function (35.1) is bounded above, and there exists an optimal solution to program (35) that we denote

$$\lambda', \pi', d', \varepsilon', \delta'. \tag{36}$$

This defines the efficient target of DMU (X_o, Y_o) as

$$(X', Y') = (X^* - \delta', Y^* + \varepsilon'). \tag{37}$$

By the conditions of model (35), $(X', Y') \in \mathcal{T}_{CRS-TO}$.

Theorem 10 (Podinovski 2007b). DMU (X', Y') in (37) is efficient in technology $\mathcal{T}_{CRS-TO}$.

Obviously, if all optimal slacks in (35), and hence the optimal value σ^*, are equal to zero, the efficient target (X', Y') coincides with the radial target (X^*, Y^*). In particular, DMU (X_o, Y_o) is efficient if and only if $(X_o, Y_o) = (X', Y')$.

In the case of VRS, model (35) requires an additional normalising condition (7). The same formula (37) defines the efficient target (X', Y') in this case.

Note that the inequality (35.4) in model (35) guarantees that the maximisation of the sum of component slacks (35.1) is performed within the technology by requiring that inputs remain nonnegative. As shown by example in Podinovski (2007b), the simple maximisation of the sum of slacks without condition (35.4) (in this case d could be assumed to be a zero vector) may result in negative values of some of the inputs.

Remark 2. Model (35) is an additive CRS DEA model based on technology $\mathcal{T}_{CRS-TO}$. It assesses the efficiency of the unit (X^*, Y^*) by maximising the sum of component slacks ε_r and δ_i, provided the resulting unit remains within the technology (and, in particular, does not have negative inputs). In the case of VRS, we need to add the normalising condition (7) to the constraints of model (35).

Model (35) and its VRS variant become standard additive DEA models (Charnes et al. 1985) in the absence of trade-offs (6). Indeed, in this case the trade-off terms on the left-hand side of conditions (35.2) and (35.3) are omitted. Furthermore, the maximisation of the sum of slack variables in (35) implies that at optimality $d = 0$, and therefore vector d can be removed from the formulation. Finally, the nonnegativity condition (35.4) is redundant because, in the absence of trade-offs, it follows from (35.3).

Like conventional additive DEA models, model (35) and its VRS variant can be used independently for the assessment of efficiency of any unit $(X^*, Y^*) \in \mathcal{T}_{CRS-TO}$, without the need to perform the first (radial projection) optimisation stage.

7.3 Stage 3: Identifying Reference Sets of Efficient Peer Units

In conventional DEA models without weight restrictions (production trade-offs), the reference set of efficient peers consists of the observed DMUs *j* such that $\lambda_j > 0$ in an optimal solution to the second-stage optimisation model. In a DEA model with weight restrictions, an observed DMU *j* with a strictly positive value λ'_j in the optimal solution (36) may be inefficient – an example of this is given in Podinovski (2007b). As proved, in this case there exists an alternative optimal solution to program (35) that results in the same efficient target (X', Y') and for which the condition $\lambda_j > 0$ implies that the observed unit *j* is efficient, for all *j*. Identifying such an optimal solution to (35) requires solving another linear program.

As with Stage 2, the computations of Stage 3 should be performed only if needed. These computations do not affect the radial efficiency, radial targets and efficient targets already obtained at Stages 1 and 2.

Following Podinovski (2007b), efficient peers of DMU (X_o, Y_o) corresponding to the efficient target (X', Y') can be obtained by maximising the sum of components of vector *d* as the secondary goal in program (35), while keeping vectors ε' and δ' at their optimum level as in (36). In this case, by (37), the constant vectors $Y^* + \varepsilon'$ and $X^* - \delta'$ on the right-hand side of conditions (35.2) and (35.3) can be replaced by Y' and X', respectively. The resulting model takes on the form:

$$D^* = \max \quad \sum_{i=1}^{m} d_i , \tag{38.1}$$

$$\text{subject to} \qquad \sum_{j=1}^{n} \lambda_j Y_j + \sum_{t=1}^{K} \pi_t Q_t = Y', \tag{38.2}$$

$$\sum_{j=1}^{n} \lambda_j X_j + \sum_{t=1}^{K} \pi_t P_t + d = X', \tag{38.3}$$

$$\lambda, \pi, d \geq 0. \tag{38.4}$$

Note that the inequality (35.4) no longer contains decision variables (because the vector $\delta = \delta'$ is kept constant) and is omitted as redundant in program (38).

Because the objective function of program (38) is bounded above, there exists an optimal solution $\tilde{\lambda}, \tilde{\pi}, \tilde{d}$ to this program. Taken together with the constant vectors ε' and δ', solution

$$\tilde{\lambda}, \tilde{\pi}, \tilde{d}, \varepsilon', \delta' \tag{39}$$

is an optimal solution to program (35). If the optimal solution (36) to program (35) is unique, then (39) is the same as (36). Otherwise, (39) is an optimal solution to (35) that additionally maximises the sum of components of vector *d* as in (38.1).

Theorem 11 (Podinovski 2007b). If $\tilde{\lambda}_j > 0$ then DMU *j* is efficient in technology $\mathcal{T}_{CRS-TO}$ and, consequently, in the smaller standard CRS technology $\mathcal{T}_{CRS} \subset \mathcal{T}_{CRS-TO}$.

An alternative model to (38) is obtained in Podinovski (2000). It has the same objective (38.1) as above maximised over the set of constraints (35.2) – (35.5), with the additional condition

$$\sum_{r=1}^{s} \varepsilon_r + \sum_{i=1}^{m} \delta_i = \sigma^*, \tag{40}$$

and keeping vectors ε and δ variable.

The difference between model (38) and the latter model is that, by solving the former, we identify the reference sets for DMU (X_o, Y_o) that are used in the composition of its specific efficient target (X', Y') which is fixed. In the latter approach, the efficient target is not fixed. The model based on condition (40) generally has alternative optima $\lambda'', \pi'', d'', \varepsilon'', \delta''$, each identifying a generally different efficient target (X'', Y'') and the corresponding reference set of efficient peers *j*.

The above results extend to the case of VRS with obvious modifications. As noted, in the case of VRS model (35) incorporates the additional normalising equality (7). The latter should also be incorporated in (38). Let

$$\hat{\lambda}, \hat{\pi}, \hat{d} \tag{41}$$

be an optimal solution to program (38) with the condition (7).

Theorem 12 (Podinovski 2007b). If $\hat{\lambda}_j > 0$ then DMU *j* is efficient in technology $\mathcal{T}_{VRS-TO}$ and, consequently, in the smaller standard VRS technology $\mathcal{T}_{VRS} \subset \mathcal{T}_{VRS-TO}$.

The above theorem implies the existence of at least one efficient DMU $j \in J$ in any technology $\mathcal{T}_{VRS-TO}$ (under the assumption that there is no free or unlimited production, as stated beforehand).

Corollary 2. In any technology $\mathcal{T}_{VRS-TO}$, there exists at least one efficient observed DMU.

Proof of Corollary 2. Because of condition (7), in solution (41) there exists a *j* such that $\hat{\lambda}_j > 0$. By Theorem 12, DMU *j* is efficient in technology $\mathcal{T}_{VRS-TO}$. ■

Note that Corollary 2 does not unconditionally extend to the case of CRS. According to Theorem 1 stated in Charnes et al. (1990), there exists at least one efficient observed DMU in the CRS technology $\mathcal{T}_{CRS-TO}$, under the condition that weight restrictions (1) are not linked. The following example shows that the same statement is generally not true in the case of linked weight restrictions.

Example 8. Consider CRS technology $\mathcal{T}_{CRS-TO}$ discussed in Example 3 and illustrated in Figure 3. The only observed unit $A = (2,1)$ is inefficient in the CRS technology induced by itself and the single linked production trade-off $(P,Q) = (1,2)$. (The latter is equivalent to the linked weight restriction $2u - v \leq 0$.) Therefore, there are no efficient observed units in technology $\mathcal{T}_{CRS-TO}$ in Figure 3. Furthermore, the output radial efficiency of *A* is equal to 0.25. Its unique efficient target is (2,4) – it is constructed entirely from the above trade-off (P,Q) applied 4 times to the origin, with no contribution from the unit *A* itself. Therefore, unit *A* has no efficient peers among observed units, and the efficient target is composed entirely from the production trade-off. Finally note that *A* is efficient in the VRS technology $\mathcal{T}_{VRS-TO}$ in Figure 3, which is consistent with Corollary 2.

8 Conclusion

In this chapter we presented the notion of production trade-offs as the dual forms of weight restrictions. We explored various theoretical, methodological and computational issues arising from the application of production trade-offs in DEA models.

Although production trade-offs are mathematically equivalent to weight restrictions, the assessment of the former is conducted in the language of possible changes to the inputs and outputs in the technology. In contrast, the assessment of weight restrictions often involves value judgements that are more managerial in nature and not directly related to the technological possibilities.

Based on the results of this chapter, the following standard workflow can be suggested for the practical implementation of production trade-offs and weight restrictions. This consists of three steps that may need to be repeated iteratively as the model is being modified by the incorporation of additional trade-offs.

1) *Construction of production trade-offs and weight restrictions*. As illustrated in Section 3, production trade-offs should represent realistic assumptions about the technology. In practice, we should be certain that all observed DMUs would be willing to accept the simultaneous changes stated by the trade-offs.
2) *Verification that the trade-offs (or weight restrictions) do not generate free or unlimited production for the given set of observed DMUs*. As discussed in Section 6, this stage is important because, if there is free or unlimited production in the technology, the results of the next stage may be inconsistent and puzzling. Alternatively, such results may appear unproblematic but still be erroneous. This stage requires either the checking of simple inequalities or, in the case of linked weight restrictions, the use of standard DEA software with the extended set of observed DMUs.
3) *Computation of efficiency, efficient targets and efficient peers*. There are three stages in the computational procedure described in Section 7. In many practical applications only the first stage would be needed and may be performed using standard DEA software. The implementation of Stages 2 and 3 would currently require the use of general linear solvers.

The use of production trade-offs in DEA models, or the use of weight restrictions obtained from such trade-offs, is interesting for a number of reasons.

First, production trade-offs allow us to specify additional information about the technology that is not otherwise captured by the observed data and standard production assumptions. This leads to *a meaningful extension* of the conventional CRS or VRS production technology and results in a better-informed model of the production process. Furthermore, this generally improves the efficiency discrimination of the model in a technologically meaningful way.

Second, the use of production trade-offs or weight restrictions based on them does not have the well-known drawback of weight restrictions assessed by other methods. The use of the latter generally leads to an uncontrolled expansion of the model of technology. In particular, the value judgements used in the construction of weight restrictions cannot generally explain the technological meaning of the expanded technology and its new efficient frontier. As a result, the radial and efficient targets of inefficient units may not be producible. The meaning of radial efficiency as the ultimate and technologically feasible improvement factor is no longer preserved. In contrast, the assessment of production trade-offs explicitly takes into account the meaning of the resulting expansion of the technology. The use of such trade-offs or weight restrictions based on them preserves the traditional meaning of efficiency.

Third, because the use of production trade-offs results in a meaningful model of production technology, the well-established notions of productivity analysis such as returns to scale, productivity change, and other can be extended to it in a straightforward fashion. In particular, the former can be explored by the generic method of reference technologies developed by Färe et al. (1985) and further explored by Podinovski (2004c). The Malmquist productivity index in models with production trade-offs was discussed in Alirezaee and Afsharian (2010).

Fourth, the clear technological meaning of production trade-offs allows us to make relatively complex statements involving several inputs and outputs in a single trade-off or weight restriction. Examples of such statements were production trade-offs (14) and (16) and the corresponding weight restrictions (15) and (17). An advantage of such complex production trade-offs is that they generally add more points to the model of production technology than simple statements, and therefore contribute to better efficiency discrimination. It is unlikely that weight restrictions (15) and (17) could be obtained using value judgements.

Fifth, production trade-offs can be used in DEA models that do not have dual multiplier forms. An example of this is the FDH technology.

Sixth, the interpretation of weight restrictions as the dual forms of production trade-offs allows us to clarify and resolve some theoretical, methodological and computational issues arising in the context of weight restrictions. For example, as discussed, the interpretation of weight restrictions in terms of production trade-offs gives a positive answer to the long-standing question of applicability of weight restrictions in the VRS technology. In Section 6 we showed that the notion of trade-offs is instrumental in understanding the infeasibility and related problems in DEA models with weight restrictions.

References

Alirezaee MR, Afsharian M. Improving the discrimination of data envelopment analysis in multiple time periods. Intl Trans Oper Res. 2010; 17(5): 667-679.

Allen R, Athanassopoulos A, Dyson RG, Thanassoulis E. Weights restrictions and value judgements in Data Envelopment Analysis: Evolution, development and future directions. Ann Oper Res. 1997; 73: 13-34.

Amado CAF, Dyson RG. Exploring the use of DEA for formative evaluation in primary diabetes care: An application to compare English practices. J Oper Res Soc. 2009; 60(11): 1469-1482.

Amado CAF, Santos SP. Challenges for performance assessment and improvement in primary health care: The case of the Portuguese health centres. Health Policy. 2009; 91(1): 43-56.

Atici KB. Using data envelopment analysis for the efficiency and elasticity evaluation of agricultural farms. Ph.D. thesis. University of Warwick, UK. 2012.

Banker RD, Charnes A, Cooper WW. Some models for estimating technical and scale inefficiencies in data envelopment analysis. Manag Sci. 1984; 30(9): 1078-1092.

Charnes A, Cooper WW, Rhodes E. Measuring the efficiency of decision making units. Eur J Oper Res. 1978; 2(6): 429-444.

Charnes A, Cooper WW, Golany B, Seiford L. Foundations of data envelopment analysis for Pareto-Koopmans efficient empirical production functions. J Econometrics. 1985; 30(1-2): 91-107.

Charnes A, Cooper WW, Wei QL, Huang ZM. Cone ratio data envelopment analysis and multi-objective programming. Int J Syst Sci. 1989; 20(7): 1099-1118.

Charnes A, Cooper WW, Huang ZM, Sun DB. Polyhedral cone-ratio DEA models with an illustrative application to large commercial banks. J Econometrics. 1990; 46(1-2): 73-91.

Cook WD, Zhu J. Context-dependent assurance regions in DEA. Oper Res. 2008; 56(1): 69-78.

Cooper WW, Ruiz JL, Sirvent I. Choices and uses of DEA weights. In: Cooper WW, Seiford LM, Zhu J, eds. Handbook on data envelopment analysis. 2nd edition. New York: Springer Science+Business Media; 2011a; p. 93-126.

Cooper WW, Seiford LM, Zhu J. Data envelopment analysis: history, models, and interpretations. In: Cooper WW, Seiford LM, Zhu J, eds. Handbook on data envelopment analysis. 2nd edition. New York: Springer Science+Business Media; 2011b; p. 1-39.

Deprins D, Simar L, Tulkens H. Measuring labor-efficiency in post offices. In: Marchand M, Pestieau P, Tulkens H, eds. The performance of public enterprises. Amsterdam: Elsevier Science Publishers; 1984; 243-267.

Dyson RD, Thanassoulis E. Reducing weight flexibility in data envelopment analysis. J Oper Res Soc. 1988; 39(6): 563-576.

Färe R, Grosskopf S, Lovell CAK. The measurement of efficiency of production. Boston: Kluwer Academic Publishers; 1985.

Førsund FR. Weight restrictions in DEA: misplaced emphasis? J Prod Anal. 2013; 40(3): 271-283.

Halme M, Korhonen P. Restricting weights in value efficiency analysis. Eur J Oper Res. 2000; 126(1): 175-188.

Jones CN, Kerrigan EC, Maciejowski JM. On polyhedral projection and parametric programming. J Optim Theory Appl. 2008; 138(2): 207-220.

Khalili M, Camanho AS, Portela MCAS, Alirezaee MR. The measurement of relative efficiency using data envelopment analysis with assurance regions that link inputs and outputs. Eur J Oper Res. 2010; 203(3): 761-770.

Pedraja-Chaparro F, Salinas-Jimenez J, Smith P. On the role of weight restrictions in data envelopment analysis. J Prod Anal. 1997; 8(2): 215-230.

Podinovski VV, Athanassopoulos AD. Assessing the relative efficiency of decision making units using DEA models with weight restrictions. J Oper Res Soc. 1998; 49(5): 500-508.

Podinovski VV. Side effects of absolute weight bounds in DEA models. Eur J Oper Res. 1999; 115(3): 583-595.

Podinovski VV. Weight restrictions and production trade-offs in DEA models. Working paper No. 330. Warwick Business School. 2000. Available at http://www.wbs.ac.uk/downloads/working_papers/330.pdf

Podinovski VV. Production trade-offs and weight restrictions in data envelopment analysis. J Oper Res Soc. 2004a; 55(12): 1311-1322.

Podinovski VV. Suitability and redundancy of non-homogeneous weight restrictions for measuring the relative efficiency in DEA. Eur J Oper Res. 2004b; 154(2): 380-395.

Podinovski VV. Efficiency and returns to scale on the “no free lunch” assumption only. J Prod Anal, 2004c; 22(3): 227-257.

Podinovski VV. The explicit role of weight bounds in models of data envelopment analysis. J Oper Res Soc. 2005; 56(12): 1408-1418.

Podinovski VV. Improving data envelopment analysis by the use of production trade-offs. J Oper Res Soc. 2007a; 58(10): 1261-1270.

Podinovski VV. Computation of efficient targets in DEA models with production trade-offs and weight restrictions. Eur J Oper Res. 2007b; 181(2): 586-591.

Podinovski VV, Bouzdine-Chameeva T. Weight restrictions and free production in data envelopment analysis. Oper Res. 2013; 61(2): 426-437.

Roll Y, Cook WD, Golany B, Controlling factor weights in data envelopment analysis. IIE Trans. 1991; 23(1): 2-9.

Santos SP, Amado CAF, Rosado JR. Formative evaluation of electricity distribution utilities using data envelopment analysis. J Oper Res Soc. 2011; 62(7): 1298-1319.

Schaffnit C, Rosen D, Paradi JC. Best practice analysis of bank branches: an application of DEA in a large Canadian bank. Eur J Oper Res. 1997; 98(2): 269-289.

Shephard RW. Semi-homogeneous production functions and scaling of production. In: Eichhorn W, Henn R, Opitz O, Shephard RW, eds. Production Theory. New York: Springer-Verlag; 1974; p. 253-285.

Thanassoulis E, Dyson RG, Foster MJ. Relative efficiency assessments using data envelopment analysis: an application to data on rates departments. J Oper Res Soc. 1987; 38(5): 397-411.

Thanassoulis E, Allen R. Simulating weights restrictions in data envelopment analysis by means of unobserved DMUs. Manag Sci. 1998; 44(4): 586-594.

Thanassoulis E, Portela MC, Allen R. Incorporating value judgements in DEA. In: Cooper WW, Seiford LM, Zhu J, eds. Handbook on data envelopment analysis. Boston: Kluwer Academic Publishers; 2004; p. 99-138.

Thanassoulis E, Portela MCS, Despić O. Data envelopment analysis: the mathematical programming approach to efficiency analysis. In: Fried HO, Lovell CAK, Schmidt SS, eds. The Measurement of productive efficiency and productivity growth. New York: Oxford University Press; 2008; p. 251-420.

Thompson RG, Langemeier LN, Lee CT, Lee E, Thrall RM. The role of multiplier bounds in efficiency analysis with application to Kansas farming. J Econometrics. 1990; 46(1-2): 93-108.

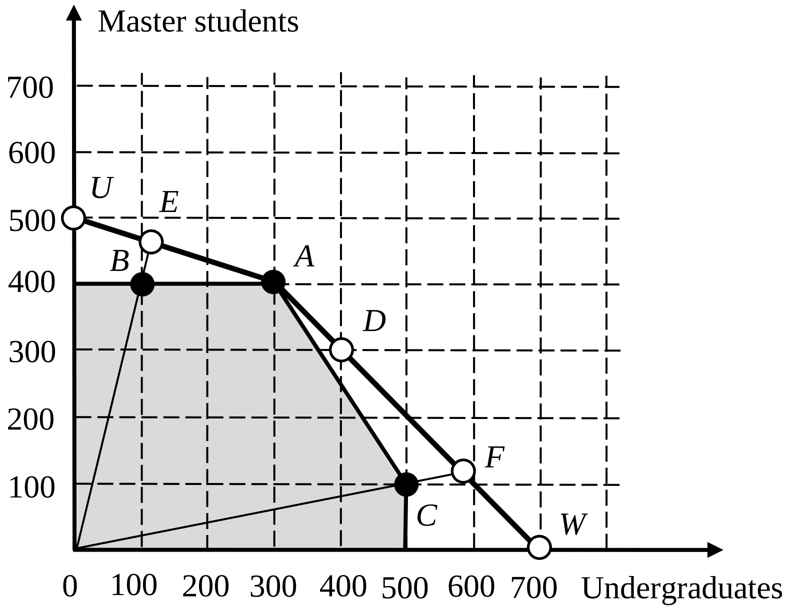

Figure 1 Production trade-offs expanding the technology in output dimensions

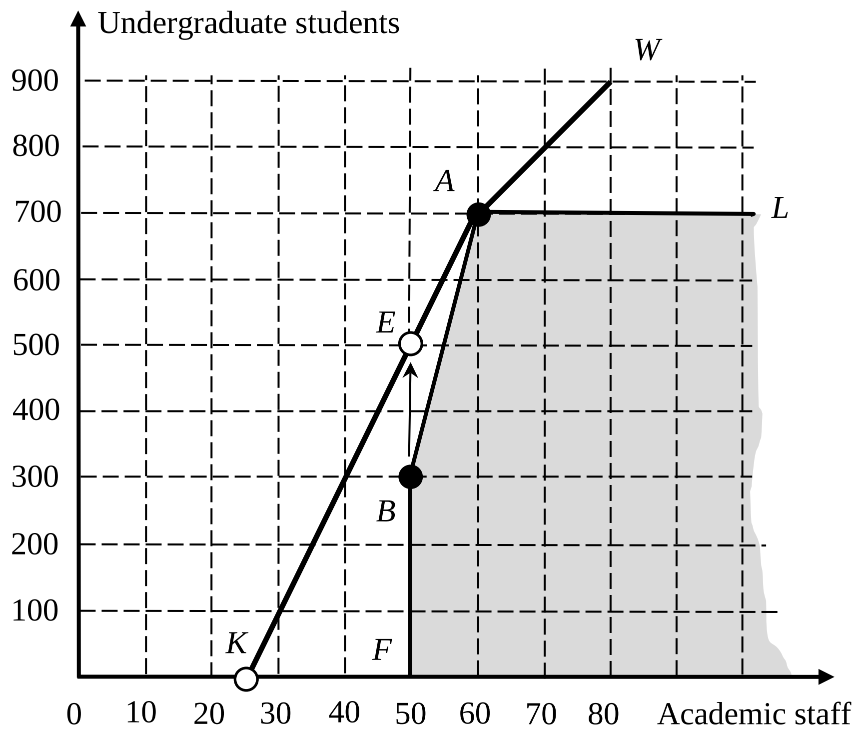

Figure 2 Linked production trade-offs expanding the VRS technology

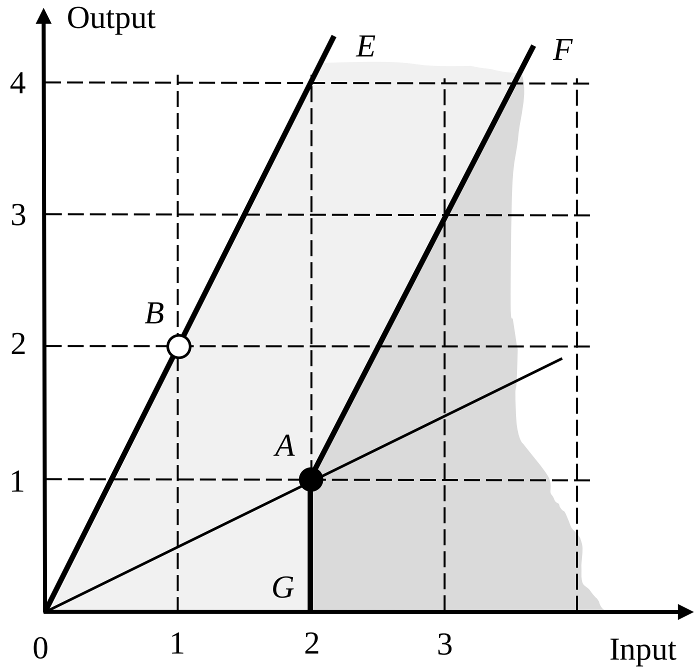

Figure 3

The VRS (dark grey) and CRS (light grey) technologies induced by unit *A* and production trade-off $(P,Q)=(1,2)$

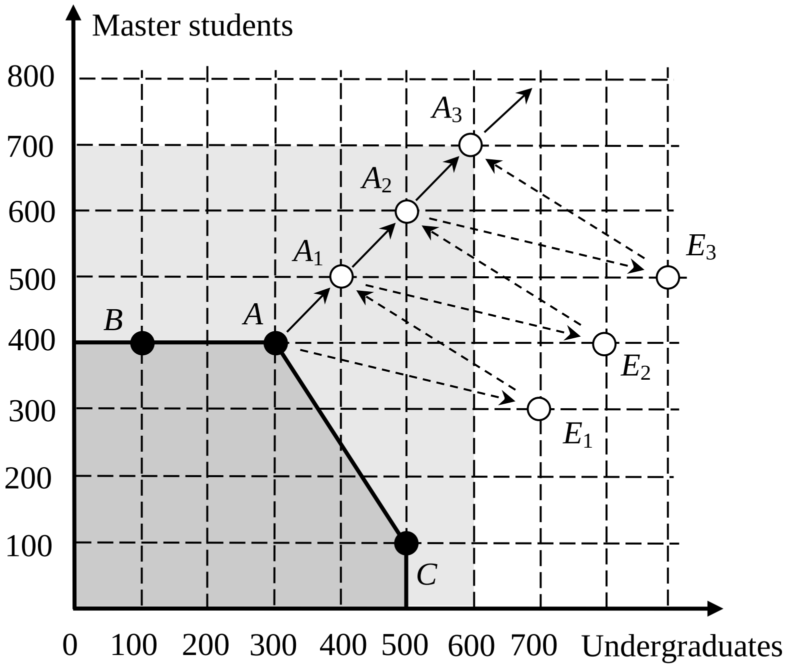

Figure 4 Free production created by two trade-offs in output dimensions

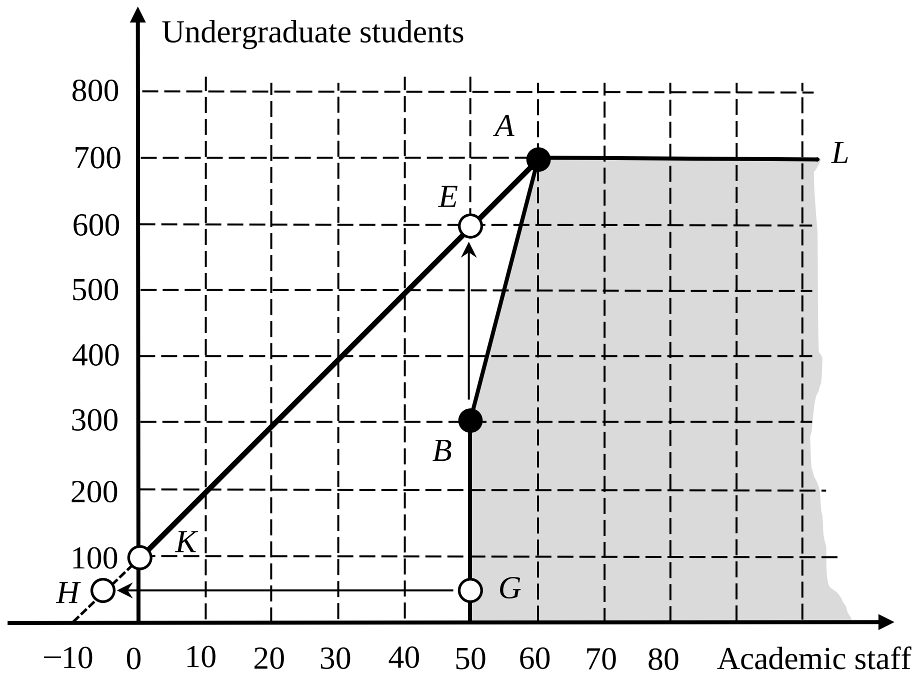

Figure 5

Free production created by the linked trade-off $(\tilde{P}, \tilde{Q}) = (-1, -10)$ in the VRS technology, and the negative "efficiency" of unit *G*.

		Departments						
		D1	D2	D3	D4	D5	D6	D7
Outputs:	Undergraduates	800	1200	1680	630	1070	1450	1550
	Master students	200	500	250	410	120	230	0
	Publications	90	21	2	97	11	109	3
Inputs:	Full academic staff	92	104	64	75	62	98	63
	Research staff	15	11	0	12	1	32	0

Table 1 University departments

Department	CRS	CRS 1	CRS 2	CRS 3	CRS 4	CRS 5	CRS 6	CRS 7	CRS 8
D1	83.27	76.82	76.82	76.47	76.47	76.47	76.47	76.47	76.47
D2	97.37	97.37	73.30	69.59	69.59	69.59	69.59	69.59	69.59
D3	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
D4	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
D5	100.00	100.00	100.00	100.00	77.61	75.92	75.92	71.78	71.78
D6	100.00	100.00	100.00	86.63	86.63	86.63	86.63	86.63	86.63
D7	100.00	100.00	100.00	100.00	85.33	84.61	84.61	83.02	83.02

Table 2 Output radial efficiency (%) of departments in the CRS models with different sets of production trade-offs/weight restrictions

Department	VRS	VRS 1	VRS 2	VRS 3	VRS 4	VRS 5	VRS 6	VRS 7	VRS 8
D1	95.68	91.09	91.09	90.07	90.07	81.52	81.52	81.52	81.52
D2	100.00	100.00	100.00	100.00	100.00	95.41	95.41	95.41	92.61
D3	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
D4	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
D5	100.00	100.00	100.00	100.00	100.00	100.00	100.00	79.66	79.66
D6	100.00	100.00	100.00	100.00	100.00	94.54	94.54	94.54	94.54
D7	100.00	100.00	100.00	100.00	99.36	99.36	97.89	87.03	87.03

Table 3 Output radial efficiency (%) of departments in the VRS models with different sets of production trade-offs/weight restrictions

Department	u_1	u_2	u_3	v_1	v_2
D1	0.0004	0	0.0073	0.0116	0.0088
D2	0	0.002	0	0.0078	0.0195
D3	0.0002	0.0025	0	0.0156	0
D4	0	0	0.0103	0.0005	0.08
D5	0.0001	0.0003	0.0762	0.0066	0.59
D6	0.0003	0	0.0051	0.0082	0.0062
D7	0.0005	0	0.0543	0.0159	0.3683

Table 4 Optimal output and input weights in the standard output-oriented multiplier CRS model

Department	u_1	u_2	u_3	v_1	v_2
D1	0.0004	0.0004	0.0068	0.0120	0.0135
D2	0.0003	0.0010	0.0047	0.0128	0.0093
D3	0.0005	0.0005	0.0088	0.0156	0.0176
D4	0.0003	0.0003	0.0067	0.0103	0.0188
D5	0.0007	0.0007	0.0142	0.0218	0.0397
D6	0.0003	0.0003	0.0049	0.0086	0.0097
D7	0.0006	0.0006	0.0124	0.0191	0.0348

Table 5 Optimal output and input weights in the final output-oriented multiplier CRS model with all eight production trade-offs/weight restrictions

Department	CRS	VRS
D1	75.64	91.09
D2	23.14	23.55
D3	65.62	66.67
D4	100.00	100.00
D5	100.00	100.00
D6	86.00	100.00
D7	100.00	100.00

Table 6 Output radial efficiency (%) of departments in the CRS and VRS models with trade-offs (25) and (26) causing free production