Nonparametric production technologies with multiple component processes

This item was submitted to Loughborough University's Institutional Repository by the/an author.


Additional Information:

- This paper was accepted for publication in the journal Operations Research and the definitive published version is available at https://doi.org/10.1287/opre.2017.1667

Metadata Record: https://dspace.lboro.ac.uk/2134/26133

Version: Accepted for publication

Publisher: © (Institute for Operations Research and Management Sciences (INFORMS))

Rights: This work is made available according to the conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence. Full details of this licence are available at: https://creativecommons.org/licenses/by-nc-nd/4.0/

Please cite the published version.
Nonparametric Production Technologies with Multiple Component Processes

Victor V. Podinovski
School of Business and Economics, Loughborough University, Loughborough LE11 3TU, United Kingdom, v.podinovski@lboro.ac.uk

Ole Bent Olesen
Department of Business and Economics, The University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark, ole@sam.sdu.dk

Cláudia S. Sarrico
ISEG Lisbon School of Economics & Management, Universidade de Lisboa, 1200-781 Lisboa, and CIPES Centre for Research in Higher Education Policies, 4450-137 Matosinhos, Portugal, cssarrico@iseg.ulisboa.pt

We develop a nonparametric methodology for assessing the efficiency of decision making units operating in a production technology with several component processes. The latter is modeled by the new multiple hybrid returns-to-scale (MHRS) technology, formally derived from an explicitly stated set of production axioms. In contrast with the existing models of data envelopment analysis (DEA), the MHRS technology allows the incorporation of component-specific and shared inputs and outputs that represent several proportional (scalable) component production processes, as well as nonproportional inputs and outputs. Our approach does not require information about the allocation of shared inputs and outputs to component processes or any assumptions about this allocation. We demonstrate the usefulness of the suggested approach in an application in the context of secondary education, and also in a Monte Carlo study based on a simulated data generating process.

Key words: data envelopment analysis; efficiency; multiple-component technology; secondary education

1. Introduction

Data envelopment analysis (DEA) is a nonparametric methodology developed for assessing the efficiency of organizational units (see, e.g., Ray 2004, Cooper et al. 2007, Thanassoulis et al. 2008). The latter are referred to as decision making units (DMUs). All DMUs are assumed to represent the same production technology generally characterized by multiple inputs and outputs. The two conventional DEA models (Charnes et al. 1978, Banker et al. 1984) are based on the assumptions that the underlying technology exhibits constant or variable returns to scale (CRS and VRS).
In both CRS and VRS models it is implicitly assumed that all inputs have a positive impact on all outputs. This premise is however often too generic. In particular, some technologies may consist of a number of component production processes, although these may not be completely separate from each other. For example, Beasley (1995) considers the assessment of the teaching and research activities of universities as two production components that use both specific and shared inputs. Cook et al. (2000) and Cook and Zhu (2006, 2011) consider operations of bank branches, where the sales of financial products and counter transactions are regarded as component processes requiring generally different, although overlapping, types of resources. Cook and Green (2004), Cook et al. (2013) and Imanirad et al. (2013) discuss examples of multiple component technologies in manufacturing plants. Cherchye et al. (2013) motivate their development by an example of a large service company with several distinct activities.

The VRS and CRS models provide no means to incorporate the above information about component production processes. The use of these models in such circumstances may result in a significant overestimation of the efficiency of DMUs. A simple example of this is given in Imanirad et al. (2013). Below we reinterpret this example in the context of assessment of school efficiency.

**Example 1.** Consider assessing the efficiency of schools $A$ and $B$ in the teaching of mathematics and languages (Table 1). The two inputs are the number of teachers in the two subjects, and the two outputs are the corresponding number of students. For simplicity, we assume that all quality factors (such as the exam attainment and socio-economic background of students) are sufficiently similar and can be ignored.

In this example we may argue that the assumption of CRS is reasonable because our model includes only the quantitative aspect of the teaching process, and also because it is a common
managerial practice to maintain a certain ratio between students and teachers (although this ratio may generally be different for different schools). Therefore, for example, a 10% increase in the number of students should be technologically feasible provided the number of teachers is increased by the same factor.

It is straightforward to verify that both schools $A$ and $B$ are efficient in the CRS model.\(^1\) Note, however, that this result is counter-intuitive. Indeed, school $A$ has 100 students per teacher in mathematics and 50 in languages. School $B$ has twice as many students per teacher in each subject: 200 in mathematics and 100 in languages. These ratios suggest that school $A$ should be regarded as only 50% efficient, but the CRS model does not identify this.

In the described situation it appears reasonable to apply the assumption of CRS separately to the two component processes: the teaching of mathematics and languages. Thus, multiplying the number of teachers and students of school $B$ in mathematics and languages by the factors 0.4 and 2, respectively, we obtain the target school $A^*$ as shown in the last row of Table 1. This school has the same number of teachers and twice the number of students in each subject, compared to school $A$. The comparison of school $A$ with $A^*$ results in the reduction of assessed efficiency of the former school to 50%, which is an intuitively correct estimate.

1.1. Motivational Scenario

The above simple example illustrates the limitations of standard DEA technologies in applications with component processes. Consider a more realistic scenario. Namely, suppose we have a number of inputs and outputs that can be classed into the following three types.

- **Component-specific proportional inputs and outputs.** Inputs and outputs of this type can be assumed to be mutually proportional. In Example 1 this includes the number of teachers and students in individual subjects.

- **Shared proportional inputs and outputs.** This type includes inputs and outputs that cannot be allocated to a particular component process, and allows a range of different interpretations. For example, we may have no information about the proportions of shared inputs and outputs.
attributed to individual processes (although in principle knowing that such proportions exist). Alternatively, any specific allocation may be meaningless or at least questionable because such inputs are used jointly (simultaneously) by all processes, in the sense explored by Cherchye et al. (2013). In many cases, shared inputs and outputs may allow a combination of the above two interpretations, as partly allocated and partly joint measures.

For example, in a school assessment with several distinct programs of study, the total expenditure on all students may be regarded as a shared proportional input. An unknown part of this input may be deemed allocated (also in unknown proportions) to the individual programs, and the remaining unknown part may be used jointly, e.g., on school administration, facilities and open days.

Similarly, the total number of students with special needs may be regarded as a shared proportional output. We assume that all school programs and the total school expenditure (the latter viewed as a shared input) contribute to this output, although in some unknown proportions.

The scaling of individual component processes (by generally unequal factors) requires that the shared proportional inputs and outputs are also scaled accordingly. In our paper we propose a treatment of such inputs and outputs that aligns their scaling with that of component-specific measures.

- Nonproportional inputs and outputs. These may be shared or not shared inputs and outputs that cannot be assumed to change proportionally in line with the first two types. In the application to secondary schools discussed in §5, we argue that the academic achievements of students on entry and exit, and their socio-economic status are examples of such measures.2

In the described scenario, also arising in the reported application below, it would be useful to have a model of technology capable of incorporating the above three different types of input and output. It is clear that the standard VRS and CRS models are not suitable for this purpose.

1.2. Existing Approaches

The growing literature on multicomponent production technologies suggests different approaches to the above problem. Cook et al. (2000) and Cook and Green (2004) consider multicomponent
CRS production technologies that have component-specific and also shared inputs. Because it is generally not possible to attribute particular proportions of the shared inputs to each component process, such proportions remain variable within the model and are defined in the way that is the most favorable for the DMU under the assessment. A variant of this approach is considered in Ding et al. (2015). Cook and Hababou (2001) apply a similar methodology to the VRS model. Cook and Zhu (2011) develop a model that allows different characterizations of returns to scale for different activities in which the DMUs are engaged.

In the above approaches, the component processes are incorporated in the multiplier model in which the technology and its properties are often insufficiently transparent. For example, it may be unclear whether the notions of returns to scale introduced via the multicomponent multiplier models are consistent with the standard definitions of scale elasticity or most productive scale size. In some cases, it may also be unclear if the implicit underlying technology is convex.

A different approach is explored by Cherchye et al. (2013, 2015, 2016) who treat component processes as different technologies that use the same joint inputs. It is assumed that the latter cannot in principle be split between the processes because each process simultaneously uses all joint inputs. Below, in §7, we show that this treatment is not applicable to the shared inputs (e.g., costs) that are allocated (in unknown proportions) to different processes. In particular, treating such shared inputs as joint inputs may lead to the benchmarks (e.g., radial projections) of inefficient DMUs being located outside the true production technology.

The above approaches assume that the technology is either CRS or VRS, and are not suitable for the incorporation of proportional and nonproportional inputs and outputs in the same model. Podinovski (2004a) develops the hybrid returns-to-scale (HRS) technology that combines characteristics of CRS with respect to a subset of inputs and outputs assumed mutually proportional, and VRS with respect to the remaining inputs and outputs. An application of this model in the context of school education is discussed in Podinovski et al. (2014). However, the HRS model cannot incorporate the information that there are multiple scalable component processes.
1.3. Contribution

In our paper we develop a new *multiple hybrid returns-to-scale (MHRS) technology* that models production processes with several scalable components and allows for the specification of shared proportional and nonproportional inputs and outputs.

Our approach does not require information or even explicit assumptions about the allocation of shared inputs and outputs between component processes. Instead, we postulate that proportional scaling of component processes with some generally different factors is technologically possible, provided all shared proportional inputs are scaled by the *maximum* of all these factors. This is a safe *worst-case* assumption that is suitable for any actual (but assumed unknown) allocation of the shared inputs between the component processes. Similarly, we postulate that all shared proportional outputs are scaled by the *minimum* of the individual scaling factors. This is also a safe worst-case assumption, regardless of the actual proportion in which individual component processes contribute to the shared outputs. These worst-case assumptions are so weak that they apply to shared inputs and outputs with a range of different interpretations as discussed in §1.1.3.

We formally derive the MHRS technology from a transparent set of production axioms, including a new axiom of multiple selective proportionality. This means that the MHRS technology includes only those DMUs that can be explained by the stated axioms and does not include any arbitrary DMUs. After the MHRS technology is described in an operational form, it is straightforward to use it in conjunction with various efficiency measures.

We illustrate the usefulness of the MHRS model using two examples. The first is an application to schools in Portugal providing three different types of program for secondary education, treated as different component processes. This application indicates that, in a typical practical setting, the MHRS model provides better discrimination on efficiency than the VRS and HRS models.

The second example uses Monte-Carlo simulation for a data generating process. In this case the true efficiency of the generated DMUs is known. This allows us to show that the MHRS model recovers the relative efficiency of DMUs better than the alternative models.
2. Production Assumptions

2.1. Standard Axioms

Let $T \subset \mathbb{R}^{m+s}_{+}$ be a production technology with $m \geq 1$ inputs $i \in I = \{1, \ldots, m\}$ and $s \geq 1$ outputs $r \in O = \{1, \ldots, s\}$. Elements $(X, Y) \in T$ are DMUs, where $X \in \mathbb{R}^m_+$ and $Y \in \mathbb{R}^s_+$ are the input and output vectors, respectively. Observed DMUs are denoted $(X_j, Y_j)$, where $j \in J = \{1, \ldots, n\}$. It is assumed that $X_j \neq 0$ and $Y_j \neq 0$, for all $j \in J$. The particular DMU under the consideration is denoted DMU$_o$ or, alternatively, $(X_o, Y_o)$.

We assume that technology $T$ satisfies the following four conventional production axioms. An additional axiom reflecting the multicomponent nature of technology $T$ is discussed below.\(^4\)

**Axiom 1 (Feasibility of Observed DMUs).** $(X_j, Y_j) \in T$ for any $j \in J$.

**Axiom 2 (Free Disposability).** $(X, Y) \in T$, $Y \geq Y' \geq 0$ and $X \leq X' \implies (X', Y') \in T$.

**Axiom 3 (Convexity).** Technology $T$ is a convex set.

**Axiom 4 (Closedness).** Technology $T$ is a closed set.

2.2. Multiple Selective Proportionality

Our goal is to define a technology that allows proportional scaling of component processes as outlined in §1.1. We achieve this by requiring that technology $T$, while satisfying the above Axioms 1–4, also satisfies an additional axiom of *multiple selective proportionality*.

In line with the discussed classification of inputs and outputs, consider the following decomposition of the sets $I$ and $O$ into the mutually disjoint subsets:

$$I = \bigcup_{k=1}^{K} I^k \cup I^S \cup I^{NP}, \quad O = \bigcup_{k=1}^{K} O^k \cup O^S \cup O^{NP}. \quad (1)$$

In the sense formally stated in Axiom 5 below, the subsets $I^k$ and $O^k$ are not empty and include *proportional* inputs and outputs specific to component processes $k \in K = \{1, \ldots, K\}$. The subsets $I^S$ and $O^S$ include *shared proportional* inputs and outputs, and $I^{NP}$ and $O^{NP}$ include *nonproportional* inputs and outputs, respectively. Any of the subsets $I^S$, $O^S$, $I^{NP}$ and $O^{NP}$ may be empty sets.
Based on (1), any DMU \((X, Y) \in T\) can be stated as
\[
\begin{pmatrix}
X \\
Y
\end{pmatrix} = \begin{pmatrix}
X^1, \ldots, X^K, X^S, X^{NP} \\
Y^1, \ldots, Y^K, Y^S, Y^{NP}
\end{pmatrix}.
\]  
(2)

In order to formalize the assumption that any component processes \(k \in K\) of DMU \((X, Y)\) can be scaled by individual multipliers \(\alpha^k\), we need to describe simultaneous changes to the shared proportional and nonproportional inputs and outputs that make this scaling technologically feasible.

Let \(\alpha = (\alpha^1, \alpha^2, \ldots, \alpha^K) \in \mathbb{R}_+^K\) be the vector of multipliers that describe proportional changes (increase or decrease) of component processes \(k \in K\). Define
\[
\alpha_{\min} = \min \{\alpha^k \mid k \in K\}, \quad \alpha_{\max} = \max \{\alpha^k \mid k \in K\}.
\]

Consider the following three scenarios.

1) Suppose there are no nonproportional inputs and outputs: \(I^{NP} = O^{NP} = \emptyset\). Assume that, for all \(k \in K\), all inputs from each set \(I^k\) are multiplied by the corresponding factor \(\alpha^k\). Because we do not generally know what proportion of the shared inputs from the set \(I^S\) is used for each of the component processes, we require that all shared inputs are multiplied by \(\alpha_{\max}\). This conservative assumption guarantees that the resulting shared inputs are sufficient for the scaling of the component outputs from the sets \(O^k\) by \(\alpha^k\). Similarly, because we do not generally know what proportion of shared outputs from the set \(O^S\) is produced by each process \(k\), we can only safely assume that the shared proportional outputs are scaled by the smallest factor \(\alpha_{\min}\).

2) Consider a more general case that includes nonproportional inputs but not outputs: \(I^{NP} \neq \emptyset\), \(O^{NP} = \emptyset\). As discussed in §1.1, such inputs may often reflect exogenous socio-economic and quality factors associated with the production process. Often such inputs have no quantifiable impact on the proportional component processes (e.g., the number of students in the school example). This situation is modeled by assuming that the scaling of component processes remains technologically possible provided the inputs from the set \(I^{NP}\) are kept unchanged.

3) Consider the general case in which the set of nonproportional outputs \(O^{NP}\) is not empty. Suppose that \(\alpha_{\min} > 1\). This implies that each process \(k\) has an increased level of resources that
should have a positive impact on all outputs, including nonproportional ones. However, because we do not know by how much the nonproportional outputs should increase, we model this by making the worst-case assumption that such outputs can at least remain constant.

Suppose that \( \alpha_{\text{min}} < 1 \) and therefore at least one input is reduced. To describe the corresponding change of nonproportional outputs from the set \( O^{NP} \), we can employ two equivalent assumptions. First, because we generally do not know by how much such outputs should be reduced, we may use the worst-case approach, i.e., assume that all outputs from the set \( O^{NP} \) are reduced to zero. Second, we may argue that in most applications we should not expect a reduction of nonproportional outputs by a factor less than \( \alpha_{\text{min}} \). Therefore, the nonproportional outputs can still be produced at least in proportion \( \alpha_{\text{min}} \), the same as all shared proportional outputs. Proposition 1 stated below shows that, in a freely disposable and convex technology, both assumptions are equivalent.

Based on the worst-case assumption about the nonproportional outputs, the described changes to DMU \((X,Y)\) result in the following scaled DMU which depends on the vector \( \alpha \):

\[
\begin{pmatrix}
X(\alpha) \\
Y(\alpha)
\end{pmatrix} = \begin{pmatrix}
\alpha^1X^1, \ldots, \alpha^KX^K, \alpha^{\text{max}}X^{S}, X^{NP} \\
\alpha^1Y^1, \ldots, \alpha^KY^K, \alpha^{\text{min}}Y^{S}, Y^{NP}(\alpha)
\end{pmatrix},
\]

where \( Y^{NP}(\alpha) = Y^{NP} \) if \( \alpha_{\text{min}} \geq 1 \), and \( Y^{NP}(\alpha) = 0 \) if \( 0 \leq \alpha_{\text{min}} < 1 \).

The following axiom formalizes the assumption that the selective scaling of any DMU \((X,Y)\) defined by formula (3) should be technologically feasible.

**Axiom 5 (Multiple Selective Proportionality).** Let \((X,Y) \in T\). Then \((X(\alpha),Y(\alpha)) \in T\), \( \forall \alpha \in \mathbb{R}_+^K \).

Let us show that in any freely disposable and convex technology \( T \), the worst-case assumption about the outputs \( Y^{NP}(\alpha) \) incorporated in formula (3) is equivalent to the assumption of gradual reduction of these outputs by \( \alpha_{\text{min}} \). Define DMU \((\hat{X}(\alpha),\hat{Y}(\alpha))\) as in formula (3) in which we replace the vector \( Y^{NP}(\alpha) \) by the vector \( \hat{Y}^{NP}(\alpha) = \min\{1, \alpha_{\text{min}}\} Y^{NP} \). This definition corresponds to the gradual, proportional reduction of outputs \( Y^{NP} \) by the factor \( \alpha_{\text{min}} \) if the latter is less than 1.

**Axiom 5*.** Let \((X,Y) \in T\). Then \((\hat{X}(\alpha),\hat{Y}(\alpha)) \in T\), \( \forall \alpha \in \mathbb{R}_+^K \).
An illustration of Axiom 5 for different vectors $\alpha = (\alpha^1, \alpha^2)$. The values marked by symbol $*$ are assumed by Axiom 5$^*$.

<table>
<thead>
<tr>
<th>DMU A</th>
<th>$\alpha = (\alpha^1, \alpha^2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1.5, 2)</td>
</tr>
<tr>
<td>Input 1</td>
<td>1</td>
</tr>
<tr>
<td>Input 2</td>
<td>1</td>
</tr>
<tr>
<td>Input 3</td>
<td>1</td>
</tr>
<tr>
<td>Input 4</td>
<td>1</td>
</tr>
<tr>
<td>Output 1</td>
<td>1</td>
</tr>
<tr>
<td>Output 2</td>
<td>1</td>
</tr>
<tr>
<td>Output 3</td>
<td>1</td>
</tr>
<tr>
<td>Output 4</td>
<td>1</td>
</tr>
</tbody>
</table>

Proposition 1. Let technology $T$ satisfy Axioms 2 and 3. Then Axiom 5 implies Axiom 5$^*$, and vice versa, i.e., these two axioms are equivalent.

The proof of this and the other statements are given in Appendices A–C.

Example 2. To illustrate Axioms 5 and 5$^*$, consider a numerical example. Table 2 shows DMU A whose four inputs and four outputs are all equal to 1. Suppose there are two component processes, i.e., $K = 2$. Let the sets $I^1$, $I^2$, $I^S$ and $I^{NP}$ contain Inputs 1, 2, 3 and 4, respectively. Similarly, let the sets $O^1$, $O^2$, $O^S$ and $O^{NP}$ contain Outputs 1, 2, 3 and 4, respectively.

Table 2 shows DMU $(X(\alpha), Y(\alpha))$ defined by formula (3), for three different vectors $\alpha = (\alpha^1, \alpha^2)$. Note that Input 3 is always changed with $\alpha^{max}$ and Output 3 with $\alpha^{min}$. If $\alpha^{min} < 1$, Output 4 is conservatively reduced to 0. However, according to Axiom 5$^*$ (equivalent to Axiom 5 in a convex and freely disposable technology), Output 4 is reduced with $\alpha^{min}$, the same as the shared Output 3.

3. The Multiple Hybrid Returns-to-Scale Technology

Below we formally develop the multiple hybrid returns-to-scale (MHRS) technology $T_{MHRS}$. Following Banker et al. (1984), technology $T_{MHRS}$ is defined using the minimum extrapolation principle, i.e., as the smallest technology that satisfies Axioms 1–5. This principle guarantees that technology $T_{MHRS}$ does not include any arbitrary DMUs that cannot be explained by the stated axioms.
3.1. A Preliminary Nonlinear Technology

To simplify interpretation, we first consider an intermediate technology defined by Axioms 1–3 and 5 only. This technology is generally not a closed set, but it is intuitively transparent and is different from technology $T_{MHRS}$ only with respect to its limit points.

**Definition 1.** Technology $T^*$ is the intersection of all technologies $T \subset \mathbb{R}_+^{m+s}$ that satisfy Axioms 1–3 and 5.

It is straightforward to verify that technology $T^*$ itself satisfies Axioms 1–3 and 5. Consequently, by Definition 1, technology $T^*$ is the *smallest* technology that satisfies these axioms. The following theorem gives its constructive definition.

**Theorem 1.** Technology $T^*$ is the set of all DMUs $(X,Y) \in \mathbb{R}_+^{m+s}$ for which there exist vectors $\lambda \in \mathbb{R}^n$ and $\alpha_j \in \mathbb{R}^K$, $j \in J$, such that the following conditions are true:

\[
\begin{align*}
\sum_{j=1}^{n} \lambda_j \left( \alpha_j^{k} X_j^k \right) & \leq X^k, & \forall k, \quad (4a) \\
\sum_{j=1}^{n} \lambda_j \left( \alpha_j^{\text{max}} X_j^S \right) & \leq X^S, \quad (4b) \\
\sum_{j=1}^{n} \lambda_j X_j^{NP} & \leq X^{NP}, \quad (4c) \\
\sum_{j=1}^{n} \lambda_j \left( \alpha_j^{k} Y_j^k \right) & \geq Y^k, & \forall k, \quad (4d) \\
\sum_{j=1}^{n} \lambda_j \left( \alpha_j^{\text{min}} Y_j^S \right) & \geq Y^S, \quad (4e) \\
\sum_{j=1}^{n} \lambda_j \left( \min\{1, \alpha_j^{\text{min}}\} Y_j^{NP} \right) & \geq Y^{NP}, \quad (4f) \\
1^T \lambda & = 1, \quad (4g) \\
\lambda, \alpha_j & \geq 0, \forall j. \quad (4h)
\end{align*}
\]

The meaning of conditions (4) is intuitively clear. Denote $(\tilde{X}, \tilde{Y})$ the DMU appearing as the result of calculations on the left-hand side of conditions (4a)–(4f). DMU $(\tilde{X}, \tilde{Y})$ is a convex combination of observed DMUs $(X_j, Y_j)$, each proportionally scaled in a selective way by the corresponding vector...
$\alpha_j = (\alpha^1_j, \ldots, \alpha^K_j)$ as described by Axiom 5* (equivalent to Axiom 5). According to our assumptions stated in Axioms 1, 3 and 5, DMU $(\tilde{X}, \tilde{Y})$ must be in technology $T^*$. The DMU $(X, Y)$ on the right-hand side is dominated by DMU $(\tilde{X}, \tilde{Y})$ and, by Axiom 2, is also in technology $T^*$.

It is straightforward to show that technology $T^*$ is generally not a closed set and therefore does not satisfy Axiom 4. 6

3.2. The MHRS technology $T_{\text{MHRS}}$

We define the MHRS technology $T_{\text{MHRS}}$ as the smallest technology that satisfies Axioms 1–5, by using the minimum extrapolation principle.

DEFINITION 2. Technology $T_{\text{MHRS}}$ is the intersection of all technologies $T \subset \mathbb{R}^{m+s}_+$ that satisfy Axioms 1–5.

To obtain a constructive statement of technology $T_{\text{MHRS}}$, we replace the variables $\alpha^k_j$ in the statement (4) of technology $T^*$ by the sign-free variables $\mu^k_j$ as follows:

$$\lambda_j \alpha^k_j = \lambda_j + \mu^k_j, \quad \forall j, k.$$  \hspace{1cm} (5)

Note that the case $\alpha^k_j > 1$ corresponds to $\mu^k_j > 0$, and $\alpha^k_j < 1$ corresponds to $\mu^k_j < 0$. The following result shows that the above transformation expands technology $T^*$ to the closed technology $T_{\text{MHRS}}$.

THEOREM 2. Technology $T_{\text{MHRS}}$ is the set of all DMUs $(X, Y) \in \mathbb{R}^{m+s}_+$ for which there exist vectors $\lambda \in \mathbb{R}^n$ and $\mu_j \in \mathbb{R}^K$, $j \in J$, such that the following conditions are true:

$$\sum_{j=1}^{n} (\lambda_j + \mu^k_j) X^k_j \leq X^k, \quad \forall k,$$  \hspace{1cm} (6a)

$$\sum_{j=1}^{n} \max_k \{\lambda_j + \mu^k_j\} X^S_j \leq X^S,$$  \hspace{1cm} (6b)

$$\sum_{j=1}^{n} \lambda_j X^N^P \leq X^N^P,$$  \hspace{1cm} (6c)

$$\sum_{j=1}^{n} (\lambda_j + \mu^k_j) Y^k_j \geq Y^k, \quad \forall k,$$  \hspace{1cm} (6d)

$$\sum_{j=1}^{n} \min_k \{\lambda_j + \mu^k_j\} Y^S_j \geq Y^S,$$  \hspace{1cm} (6e)
The following result restates conditions (6) in a linearized form.

**Proposition 2.** Technology $T_{MHRS}$ is equivalently stated as the set of all DMUs $(X,Y) \in \mathbb{R}^{m+s}_+$ for which there exist vectors $\lambda, \rho, \sigma, \tau \in \mathbb{R}^n$ and $\mu_j \in \mathbb{R}^K$, $j \in J$, such that

\[
\sum_{j=1}^n (\lambda_j + \mu_j^k) X_j^k \leq X^k, \quad \forall k, \quad (7a)
\]

\[
\sum_{j=1}^n (\lambda_j + \rho_j) X_j^S \leq X^S, \quad (7b)
\]

\[
\sum_{j=1}^n \lambda_j X_j^{NP} \leq X^{NP}, \quad (7c)
\]

\[
\sum_{j=1}^n (\lambda_j + \mu_j^k) Y_j^k \geq Y^k, \quad \forall k, \quad (7d)
\]

\[
\sum_{j=1}^n (\lambda_j + \sigma_j) Y_j^S \geq Y^S, \quad (7e)
\]

\[
\sum_{j=1}^n (\lambda_j - \tau_j) Y_j^{NP} \geq Y^{NP}, \quad (7f)
\]

\[1^T \lambda = 1, \quad (7g)\]

\[- \lambda_j + \tau_j \leq 0, \quad \forall j, \quad (7h)\]

\[- \mu_j^k + \tau_j \leq 0, \quad \forall j, k, \quad (7i)\]

\[\mu_j^k - \rho_j \leq 0, \quad \forall j, k, \quad (7j)\]

\[- \mu_j^k + \sigma_j \leq 0, \quad \forall j, k, \quad (7k)\]

\[\lambda, \tau \geq 0, \ \rho, \sigma, \mu_j \text{ sign free vectors, } \forall j. \quad (7l)\]

**Proposition 3.** Technology $T_{MHRS}$ is a polyhedral set. If $I^{NP} = O^{NP} = \emptyset$, technology $T_{MHRS}$ is a polyhedral cone.\(^7\)
Proposition 4. Technology $T_{MHRS}$ is the closure of technology $T^*$, i.e., $T_{MHRS} = \text{cl } T^*$.

Remark 1. As follows from Proposition 4 and the above discussion of the meaning of conditions (4), technology $T_{MHRS}$ consists of the following (not mutually exclusive) types of DMUs: observed DMUs $(X_j, Y_j)$, $j \in J$, scaled according to Axiom 5* by the vectors $\alpha_j = (\alpha_{j1}, \ldots, \alpha_{jK})$ calculated from equalities (5), their convex combinations, DMUs dominated by such convex combinations, and all limit points (DMUs) of the previous types.$^8$

Remark 2. If there is a single production process and there are no nonproportional inputs and outputs, technology $T_{MHRS}$ is the conventional CRS technology. Indeed, in this case, $K = 1$ and $I^S = O^S = I^{NP} = O^{NP} = \emptyset$. Consider the statement of technology $T_{MHRS}$ by conditions (6). In the described scenario, constraints (6b), (6c), (6e) and (6f) are removed. The normalizing equality (6g) becomes redundant and is also omitted. Removing the superscript $k$ and denoting $\hat{\lambda}_j = \lambda_j + \mu_j$, we obtain the standard statement of the CRS technology. Furthermore, if there is a single production process and at least one nonproportional input or output, technology $T_{MHRS}$ is the HRS technology (Podinovski 2004a).$^9$

Remark 3. In some applications based on technology $T_{MHRS}$, it may be useful to define its closed cone extension $T_{C-MHRS}$. Examples include the evaluation of scale efficiency of DMUs and its use in the decomposition of the Malmquist index of productivity change. In all such applications the cone technology $T_{C-MHRS}$ plays the role of a reference technology for the true technology $T_{MHRS}$, in the same way as the standard CRS technology is used as a reference technology for the VRS technology (Färe et al. 1994, Grosskopf 2003). Similar to the proofs of Theorem 2 and Proposition 2, it is straightforward to show that technology $T_{C-MHRS}$ is defined by the conditions (6) or (7) from which the normalizing equality $1^T \lambda = 1$ is omitted.

4. MHRS DEA Models

The description of technology $T_{MHRS}$ by linear expressions (7) makes the task of efficiency assessment of any DMU in this technology straightforward. This includes both radial and nonradial efficiency measures, e.g., those based on directional distance functions (Chambers et al. 1998).
4.1. Envelopment Models

To be specific, we consider the assessment of output radial efficiency of DMU\(_o\). By Proposition 2, this is obtained by taking the inverse of the optimal value \(\eta^*\) of the following linear program:

\[
\eta^* = \max \eta \\
\text{s.t. } \sum_{j=1}^{n} (\lambda_j + \mu_j^k) X_j^k \leq X_o^k, \quad \forall k, \\
\sum_{j=1}^{n} (\lambda_j + \rho_j) X_j^S \leq X_o^S, \\
\sum_{j=1}^{n} \lambda_j X_j^{NP} \leq X_o^{NP}, \\
- \sum_{j=1}^{n} (\lambda_j + \mu_j^k) Y_j^k + \eta Y_o^k \leq 0, \quad \forall k, \\
- \sum_{j=1}^{n} (\lambda_j + \sigma_j) Y_j^S + \eta Y_o^S \leq 0, \\
- \sum_{j=1}^{n} (\lambda_j - \tau_j) Y_j^{NP} + \eta Y_o^{NP} \leq 0, \\
\eta \text{ sign free.}
\]

The above model identifies the output radial projection of DMU\(_o\) on the boundary of the MHRS technology. Because the radial target \((\eta^* Y_o, X_o)\) is generally only weakly efficient, the usual second optimization stage maximizing the sum of component slacks needs to be performed in order to test for possible mix inefficiency (Cooper et al. 2007).

4.2. The Dual Interpretation

The standard dual to the envelopment model (8) is presented in Appendix C. Its interpretation becomes clear after a rearrangement of its terms that leads to the following nonlinear formulation.

**Theorem 3.** The dual to the output-oriented program (8) is equivalently stated as follows:

\[
\begin{align*}
\min & \quad v^\top X_o + \omega \\
\text{s.t.} & \quad u^\top Y_o = 1, \\
& \quad v^\top X_j - u^\top Y_j + \omega \geq 0, \quad \forall j,
\end{align*}
\]


The interpretation of the above multiplier model is consistent with the standard interpretation of the VRS and CRS models. The objective function (9a) and constraints (9b) and (9c) are the same as in the output-oriented VRS model. In particular, conditions (9c) imply that DMU$_o$ is benchmarked against all observed DMUs in the sense explored by Charnes et al. (1978) and Banker et al. (1984). Moreover, as we remark below, the sign-free variable $\omega$ plays the same role in the determination of the type of returns to scale as in the standard VRS model.

DMU$_o$ is also benchmarked against two further groups of DMUs that appear in inequalities (9d) and (9e). In particular, inequalities (9d) mean that DMU$_o$ is benchmarked against all proportional component processes $k \in K$ of all observed DMUs $(X_j, Y_j), j \in J$. In these inequalities, the multipliers $p^k_j$ and $q^k_j$ are interpretable as the fractions of vectors of shared proportional inputs $X^S_j$ and outputs $Y^S_j$ attributable to component process $k$. Each component process $k \in K$ of DMU $j \in J$ in inequality (9d) uses the vector of proportional inputs $(X^k_j, p^k_j X^S_j)$, to produce the vector of proportional outputs $(Y^k_j, q^k_j Y^S_j)$, where the subvectors $X^k_j, X^S_j, Y^k_j$ and $Y^S_j$ are as defined in (2).

The inputs and outputs of these subvectors are multiplied by the corresponding weights $v_i$ and $u_r$.

The inequalities (9e) correspond to the $n$ inactive DMUs (Kuosmanen 2005). Each inactive DMU $j$ is obtained by scaling the observed DMU $j$ with the zero vector $\alpha$ as in Axiom 5. All inputs and outputs of an inactive DMU $j$ are equal to zero, except for the vector $X^{NP}$ of nonproportional inputs $i \in I^{NP}$. Although inactive DMUs may be viewed as pure theoretical constructs, inequalities (9e) imply that DMU$_o$ is benchmarked against observed DMUs $j$ scaled down using vectors $\alpha$ with very small positive components $\alpha^k$.

All of the above DMUs against which DMU$_o$ is benchmarked are in technology $T_{MHRS}$ and therefore assumed producible. Any optimal solution to program (9) shows DMU$_o$ in the best light...
in comparison to these units. This includes the most favorable vectors of input and output weights \( v \) and \( u \), and also the proportions \( p^k_j \) and \( q^k_j \) in which the shared vectors \( X^S_j \) and \( Y^S_j \) of each DMU \( j \) are divided between the processes \( k \in K \).

**Remark 4.** It is interesting to note that Axioms 1–5 do not postulate any apportioning mechanism for the shared inputs and outputs between component processes. However, in the multiplier model (9) that formally follows from these axioms, the fractions of shared inputs and outputs \( p^k_j \) and \( q^k_j \) depend on the observed DMU \( j \) and component \( k \), but do not depend on the input \( i \in I^S \) or output \( r \in O^S \). This allocation mechanism arises from the worst-case scenario embedded in the treatment of shared inputs and outputs by Axiom 5. It is different from the approach of Cook et al. (2000) and related work in which the above fractions (for inputs only) are postulated to depend (in our notation) on the process \( k \) and input \( i \in I^S \) but are assumed to be equal for all observed DMUs \( j \).

**Remark 5.** A possible extension to the MHRS model is the incorporation of restrictions on proportions \( p^k_j \) and \( q^k_j \) in which the shared inputs and outputs can be attributed to the production processes \( k \), similar to the approach explored in Cook et al. (2000). In our case, in order to avoid nonlinear formulations, equivalent restrictions can be incorporated in the linear programming analogue of program (9) shown in Appendix C. However, any such restrictions on proportions would, by duality, modify the envelopment model (8) and require a reinterpretation of the underlying technology or the meaning of efficiency measure. This is similar to the use of weight restrictions in DEA models (see, e.g., Allen et al. 1997, Podinovski 1999, 2004b).

**Remark 6.** The dual program (9) can be used for the returns-to-scale (RTS) characterization of efficient DMUs in technology \( T_{MHRS} \). As follows from a more general result established by Podinovski et al. (2016)\(^{11}\), the one-sided (left-hand and right-hand) scale elasticities and the type of RTS in the MHRS technology are defined by the extreme optimal values \( \omega_{\text{min}} \) and \( \omega_{\text{max}} \) of the variable \( \omega \) in program (9). A similar relationship is well known for the standard VRS model (Banker

Remark 7. In Remark 2 we noted that the MHRS technology with a single production process (and hence with no shared inputs and outputs) and with no nonproportional inputs and outputs is the standard CRS technology. The same fact is observed from the multiplier model (9). Indeed, in the described case, the inequalities (9e) and (9f) are omitted, and the inequalities (9d) are restated as $v^\top X_j - u^\top Y_j \geq 0$, for all $j \in J$. In any optimal solution to the resulting program we have $\omega = 0$. Therefore, the variable $\omega$ is redundant and can be omitted. The resulting program is the standard output-oriented multiplier CRS model.

5. An Application to Secondary Schools

Below we demonstrate the usefulness of the MHRS model by applying it to the assessment of efficiency of public schools in Portugal providing secondary education. Our computations show that the MHRS model, while being realistic in its assumptions, provides better discrimination on efficiency than the existing VRS and HRS models.

In Portugal, school education is compulsory for all children of ages from 6 to 18. This is divided into the first 9 years of basic education and the last 3 years (referred to as years 10-12) of secondary education. At the end of basic education, all students take national exams in the Portuguese language and mathematics. All students subsequently proceed to the stage of secondary education that includes several types of program. In our study we consider the two main types: academic (preparing students for entry to higher education) and vocational (focusing on work-related skills). We also consider the third type of alternative programs. This is a collective term used for a variety of smaller, e.g., second-chance, courses that still attract a significant number of students.

At the end of secondary education students take part in national exams. Exam scores are used as admissions criteria for entry to higher education institutions.
Table 3 Descriptive statistics for 505 schools in the application to schools.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Mean</th>
<th>Median</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Program-specific proportional inputs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Input 1: Classes for academic programs</td>
<td>13.8</td>
<td>12</td>
<td>0</td>
<td>60</td>
<td>10.25</td>
</tr>
<tr>
<td>Input 2: Classes for vocational programs</td>
<td>7.3</td>
<td>6</td>
<td>0</td>
<td>31</td>
<td>5.97</td>
</tr>
<tr>
<td>Input 3: Classes for alternative programs</td>
<td>0.6</td>
<td>0</td>
<td>0</td>
<td>14</td>
<td>1.66</td>
</tr>
<tr>
<td><strong>Shared proportional input</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Input 4: Expenditure (Million Euro)</td>
<td>2.23</td>
<td>2.02</td>
<td>0.04</td>
<td>8.43</td>
<td>1.44</td>
</tr>
<tr>
<td><strong>Nonproportional inputs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Input 5: Good numeracy on entry</td>
<td>9.8</td>
<td>6</td>
<td>0</td>
<td>80</td>
<td>11.58</td>
</tr>
<tr>
<td>Input 6: Good literacy on entry</td>
<td>16.37</td>
<td>11</td>
<td>0</td>
<td>139</td>
<td>17.71</td>
</tr>
<tr>
<td>Input 7: High socioeconomic status</td>
<td>340.27</td>
<td>267</td>
<td>4</td>
<td>1519</td>
<td>280.6</td>
</tr>
<tr>
<td><strong>Program-specific proportional outputs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Output 1: Students on academic programs</td>
<td>328.96</td>
<td>254</td>
<td>0</td>
<td>1553</td>
<td>273.45</td>
</tr>
<tr>
<td>Output 2: Students on vocational programs</td>
<td>127.95</td>
<td>101</td>
<td>0</td>
<td>597</td>
<td>108.49</td>
</tr>
<tr>
<td>Output 3: Students on alternative programs</td>
<td>11.97</td>
<td>0</td>
<td>0</td>
<td>398</td>
<td>33.29</td>
</tr>
<tr>
<td><strong>Shared proportional output</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Output 4: Students passing the year</td>
<td>385.13</td>
<td>329</td>
<td>6</td>
<td>1599</td>
<td>285.28</td>
</tr>
<tr>
<td><strong>Nonproportional outputs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Output 5: Good numeracy on exit</td>
<td>9.04</td>
<td>5</td>
<td>0</td>
<td>90</td>
<td>10.98</td>
</tr>
<tr>
<td>Output 6: Good literacy on exit</td>
<td>17.04</td>
<td>11</td>
<td>0</td>
<td>156</td>
<td>19</td>
</tr>
</tbody>
</table>

5.1. Data

We use data collected in the academic year 2012/13. In this year there were 538 schools providing secondary education and located in Continental Portugal. For homogeneity reasons, and after the data verification, we limit our sample to 505 schools. Table 3 shows the 7 inputs and 6 outputs included in the model, and the corresponding descriptive statistics.

Inputs 1–3 represent the number of classes (groups of students) taught by a school in years 10–12, separately for the three different types of program (or, for brevity, three programs). An alternative to the number of classes would be the teaching hours for each program, but these data were not available to us. It is worth noting that not all schools provide all three programs. For example, only 130 schools provide alternative programs, and only 121 schools provide all three programs.

Input 4 represents the total expenditure by a school on secondary education, from which, in order to avoid double counting of the teaching provision already represented by Inputs 1–3, we
have subtracted the teachers’ salaries. Inputs 5 and 6 show the number of students in year 12 whose results in the mathematics and language exams at the end of their basic education (i.e., taken three years before) were in the upper quartile of the national results. Input 7 reflects the socioeconomic status of the students. It is calculated by subtracting the number of students whose families obtain financial help from the state, from the total number of students.

Outputs 1–3 are the number of students in years 10–12, on each of the three types of program. Output 4 is an important metric showing the number of students who have successfully passed the year (years 10 and 11) or completed their secondary education (year 12). Outputs 5 and 6 represent students in year 12 whose exam results at the end of secondary education are in the upper quartile of the national results. Note that Inputs 5 and 6 control for the academic achievements of exactly the same students, evaluated at their entry to secondary education three years before.

5.2. Specifying the MHRS Model

Below we show that the educational process employed by the schools can be modeled by the MHRS technology with three component processes corresponding to the three types of program.

It appears reasonable to assume that the number of students on each program (Outputs 1–3) should be in proportion to the teaching provision, i.e., to the number of classes on each program (Inputs 1–3). For example, increasing the number of classes on any of the three programs by, e.g., 50% (i.e., by $\alpha = 1.5$) should enable the school to teach 50% more students on this program.

We consider the total expenditure of the school (Input 4) as a shared proportional input allocated between the three programs. Similarly, we consider the total number of students who successfully pass the year (Output 4) as a shared proportional output. Although we do not know the proportions of Input 4 and Output 4 attributable to the three programs, we can still make the most conservative assumption, as in Axiom 5. Namely, it is reasonable to accept that the three programs can scale their classes and students by the factors $\alpha_1, \alpha_2$ and $\alpha_3 > 0$, provided the school expenditure (Input 4) is scaled by the maximum of these three factors, and the number of students successfully passing the year (Output 4) is scaled by the minimum of these three factors.
Table 4  Summary statistics for the output radial efficiency in different models.

<table>
<thead>
<tr>
<th>Model</th>
<th>Number of efficient schools</th>
<th>Minimum efficiency</th>
<th>Mean efficiency</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>VRS</td>
<td>248</td>
<td>0.5441</td>
<td>0.9383</td>
<td>0.0874</td>
</tr>
<tr>
<td>CRS</td>
<td>141</td>
<td>0.4968</td>
<td>0.9049</td>
<td>0.0965</td>
</tr>
<tr>
<td>HRS</td>
<td>134</td>
<td>0.4022</td>
<td>0.8814</td>
<td>0.113</td>
</tr>
<tr>
<td>MHRS</td>
<td>120</td>
<td>0.3909</td>
<td>0.8763</td>
<td>0.1141</td>
</tr>
</tbody>
</table>

The above inputs and outputs can be seen as measures that mainly represent the teaching volume, as opposed to the teaching quality. The above assumptions treat these volume measures as being mutually proportional, disregarding the quality factors discussed next.

The quality of education is represented by Outputs 5 and 6, measuring the number of students with good results in the literacy and numeracy in the final exams. These outputs should positively depend on the student results on entry to secondary education, and also on the socioeconomic status of their families, represented by Inputs 5–7. In contrast with the volume inputs and outputs that can be assumed proportional, it is unlikely that there is a simple linear relationship between the three quality-related inputs and the two quality outputs. Therefore we exclude Inputs 5–7 and Outputs 5 and 6 from the proportional relationship and treat them as nonproportional measures.

In summary, we define the MHRS model by specifying the following sets: $I^k = \{\text{Input } k\}$, $O^k = \{\text{Output } k\}$, $k = 1, 2, 3$, $I^S = \{\text{Input } 4\}$, $O^S = \{\text{Output } 4\}$, $I^{NP} = \{\text{Inputs } 5, 6, 7\}$, and $O^{NP} = \{\text{Outputs } 5, 7\}$.

5.3. Computational Results and Discussion

We evaluate the output radial efficiency of all schools by solving the MHRS program (8). For comparison, we also solve the output-oriented VRS, CRS and HRS models. Because the CRS model is based on the assumption that all inputs and outputs are proportional which is not satisfied in the described application, the practical use of this model would be unsubstantiated. However, we use the CRS model to illustrate the theoretical relationship between different technologies.

Table 4 shows a summary of computational results. These reflect the known embedding $T_{VRS} \subseteq T_{CRS}$, and the following embedding that is straightforward to verify:
where $T_{\text{VRS}}, T_{\text{CRS}}, T_{\text{HRS}}$ and $T_{\text{MHRS}}$ are the VRS, CRS, HRS and MHRS technologies.

Reflecting embedding (10), the efficiency scores are the highest in the VRS model and the lowest in the MHRS model, for each school in our sample. The VRS model is the most conservative of the above models. It ignores the information that the three component processes are scalable, which results in the lowest discrimination on efficiency. The HRS model uses this information but does not recognise the possibility of individual scaling of the component processes. The MHRS model fully utilizes this information and provides the best discrimination on efficiency among the three theoretically acceptable models.

Neither the CRS nor the MHRS technology is a subset of each other. In our example, the efficiency of 331 schools in the MHRS model is lower than in the CRS model, for 61 schools we have the opposite result, and for 113 schools the efficiencies are equal. Note that we do not propose to use the CRS model, because of the assumed nonproportional nature of quality-related inputs and outputs. However, even if we assumed that these measures were proportional and used the CRS model, the resulting discrimination would actually be worse than of the MHRS model.

6. Experiments with Simulated Data

In this section we present results from several Monte Carlo experiments. In contrast with the above school application, the DMUs in these experiments are generated by a known data generating process (DGP), and the true efficiency of each DMU is known. This allows us to show how well the MHRS model recovers the true efficiency, compared to the VRS, HRS and CRS models.

6.1. Design of the Experiments

The DGP described below is aligned with the Portuguese school application, but some inputs and outputs are different. We assume that there are two programs of study (for example, academic and vocational, or science and arts), referred to as programs 1 and 2. We also introduce a shared proportional output that represents the number of students with special needs on both programs,
and we assume that we have no information about the split of this output by the program. These students require extra teaching hours and funding.

Each DMU (school) \( j \) is randomly generated as the vector

\[
(X_j, Y_j) = (x_j^1, x_j^2, x_j^S, x_j^{NP}, y_j^1, y_j^2, y_j^S, y_j^{NP}),
\]

where

\( x_j^1, x_j^2 \) are the teaching hours provided to all students on programs 1 and 2, including additional hours for students with special needs;

\( x_j^S \) is the total expenditure of the school, including expenditure on students with special needs;

\( x_j^{NP} \) is the total number of students with good exam results on entry to both programs;

\( y_j^1, y_j^2 \) are the numbers of students on programs 1 and 2, including students with special needs;

\( y_j^S \) is the total number of students with special needs on both programs;

\( y_j^{NP} \) is the total number of students achieving good exam results at the end of the program.

A detailed description of the above DGP is given in Appendix D. As follows from this description, DMUs \( (X_j, Y_j) \) in (11) are random members of the theoretically defined technology in which the teaching hours \( x_j^1 \) and \( x_j^2 \), and the number of students \( y_j^1 \) and \( y_j^2 \) are program-specific proportional inputs and outputs. The total expenditure \( x_j^S \) and the number of students with special needs \( y_j^S \) are the shared proportional input and output, respectively. The number of students with good results on entry and exit \( x_j^{NP} \) and \( y_j^{NP} \) are the nonproportional input and output, respectively.

In order to generate each DMU \( (X_j, Y_j) \) in (11), we first employ a more detailed process in which we know the allocation of the school expenditure and students with special needs to the two programs. For example, instead of generating the total expenditure \( x_j^S \), we first randomly generate two component expenditures \( x_j^{S1} \) and \( x_j^{S2} \) representing the split of expenditure \( x_j^S \) between the two programs. We additionally generate the random shares of expenditures \( x_j^{S1} \) and \( x_j^{S2} \) allocated to the standard teaching processes, and to students with special needs. We generate similar random shares of the teaching hours \( x_j^1 \) and \( x_j^2 \) allocated to the regular teaching and to students with special needs.
The described process is equivalent to generating 9 component inputs. We assume specific parametric production functions (shown in Appendix D), and use them to calculate 5 outputs, including the total number of students and the number of students with special needs, for both programs.

Next, we generate a random radial output efficiency score \( \theta_j \in [0, 1] \), and multiply all generated outputs by \( \theta_j \leq 1 \). We perform several series of experiments in which \( \theta_j \) is generated from the uniform distribution \( U[a, 1] \) with support \([a, 1]\), where \( a \in \{0.25, 0.5, 0.75, 0.9\} \), and from the double truncated normal distribution \( N_{[0,1]}(1, \sigma^2) \) truncated to \([0, 1]\), where \( \sigma \in \{0.04, 0.12, 0.24, 0.4\} \).

Finally, we aggregate the above component inputs and outputs, to obtain DMU \((X_j, Y_j)\) in (11). For example, we add components \( x_{j1}^{S1} \) and \( x_{j2}^{S2} \) of shared costs to obtain the single value \( x_j^S \). We also add the number of students with special needs on both programs to obtain the single value \( y_j^S \).

The described aggregation of the school expenditure and students with special needs leads to a deliberate loss of information about their split between the two programs. This simulates a practical scenario in which such split is unknown to the analyst.

### 6.2. Computational Results

We ran computational experiments assuming different distributions of the school efficiency \( \theta_j \), as described above. To illustrate the impact of the sample size, we also varied the number of schools in the sample \( n \in \{50, 100, 200, 500\} \). For each combination of the sample size \( n \) and the distribution of efficiency, we generated 100 data sets (each of the size \( n \)), and for each of them we evaluated the efficiency of the \( n \) schools in the sample using the VRS, CRS, HRS and MHRS models.

Table 5 shows the results of the Monte Carlo analysis in which the true efficiency \( \theta_j \) is generated from the uniform distribution \( U[a, 1] \) with \( a \in \{0.25, 0.5, 0.75, 0.9\} \), and for different sample sizes \( n \). Following Gong and Sickles (1992) and Ruggiero (1999), we use rank correlation as a measure of performance of the model reflecting its ability to recover the true relative efficiency scores.

The first two columns of Table 5 indicate the distribution from which the true school efficiency \( \theta_j \) is generated, and the sample size \( n \). The remaining columns show the average rank correlation between the estimated and the true efficiency scores based on the 100 replications, and the
Table 5  The average rank correlation coefficients (taken over 100 replications) between the true and estimated efficiency scores in different models. Note: The true efficiency is assumed to be uniformly distributed with $U[a, 1]$, where $a \in \{0.9, 0.75, 0.5, 0.25\}$. Each replication is based on a sample of $n \in \{50, 100, 200, 500\}$ DMUs. The standard deviations of the rank correlation coefficients are shown in parentheses. The term “Indet.” means “indeterminate”. 13

<table>
<thead>
<tr>
<th>Efficiency</th>
<th>$n$</th>
<th>VRS</th>
<th>CRS</th>
<th>HRS</th>
<th>MHRS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$U[0.9, 1]$</td>
<td>50</td>
<td>Indet.</td>
<td>(Indet.)</td>
<td>0.3149</td>
<td>(0.103)</td>
</tr>
<tr>
<td>StDev ≈ 0.03</td>
<td>100</td>
<td>Indet.</td>
<td>(Indet.)</td>
<td>0.4118</td>
<td>(0.07363)</td>
</tr>
<tr>
<td>200</td>
<td>0.4187</td>
<td>(0.04904)</td>
<td>0.4954</td>
<td>(0.05264)</td>
<td>0.5277</td>
</tr>
<tr>
<td>500</td>
<td>0.5454</td>
<td>(0.03045)</td>
<td>0.6021</td>
<td>(0.03077)</td>
<td>0.6438</td>
</tr>
<tr>
<td>$U[0.75, 1]$</td>
<td>50</td>
<td>0.4556</td>
<td>(0.1065)</td>
<td>0.5545</td>
<td>(0.09401)</td>
</tr>
<tr>
<td>StDev ≈ 0.07</td>
<td>100</td>
<td>0.543</td>
<td>(0.06526)</td>
<td>0.625</td>
<td>(0.06292)</td>
</tr>
<tr>
<td>200</td>
<td>0.6315</td>
<td>(0.0529)</td>
<td>0.6976</td>
<td>(0.0455)</td>
<td>0.7576</td>
</tr>
<tr>
<td>500</td>
<td>0.7322</td>
<td>(0.0252)</td>
<td>0.7815</td>
<td>(0.02334)</td>
<td>0.8365</td>
</tr>
<tr>
<td>$U[0.5, 1]$</td>
<td>50</td>
<td>0.619</td>
<td>(0.08636)</td>
<td>0.6935</td>
<td>(0.08689)</td>
</tr>
<tr>
<td>StDev ≈ 0.14</td>
<td>100</td>
<td>0.6836</td>
<td>(0.05972)</td>
<td>0.7513</td>
<td>(0.05875)</td>
</tr>
<tr>
<td>200</td>
<td>0.7515</td>
<td>(0.04554)</td>
<td>0.8052</td>
<td>(0.04046)</td>
<td>0.8798</td>
</tr>
<tr>
<td>500</td>
<td>0.8247</td>
<td>(0.02111)</td>
<td>0.8608</td>
<td>(0.01962)</td>
<td>0.9234</td>
</tr>
<tr>
<td>$U[0.25, 1]$</td>
<td>50</td>
<td>0.6737</td>
<td>(0.0835)</td>
<td>0.7556</td>
<td>(0.07712)</td>
</tr>
<tr>
<td>StDev ≈ 0.22</td>
<td>100</td>
<td>0.7343</td>
<td>(0.05809)</td>
<td>0.8033</td>
<td>(0.05581)</td>
</tr>
<tr>
<td>200</td>
<td>0.7888</td>
<td>(0.0458)</td>
<td>0.8428</td>
<td>(0.03707)</td>
<td>0.9252</td>
</tr>
<tr>
<td>500</td>
<td>0.8447</td>
<td>(0.02115)</td>
<td>0.8799</td>
<td>(0.01799)</td>
<td>0.9544</td>
</tr>
</tbody>
</table>

corresponding standard deviation of the rank correlation (in parentheses). As expected, the rank correlation increases with the increase of the sample size $n$, for all four models.

Table 5 shows that the rank correlation increases over the sequence of models: VRS, CRS, HRS and MHRS. These results show very clearly that the conventional VRS and CRS models perform rather poorly compared to the MHRS model. The rank correlation in the MHRS model is at least 13% higher compared to the VRS model, and at least 8% higher compared to the CRS model. The noted advantage of the MHRS model is even more pronounced for smaller sample sizes.

Note that the standard deviations shown in Table 5 are relatively low, which indicates small dispersion around the average rank correlation. However, the standard deviations almost always (with the exception of the VRS model with $n = 200$ and $n = 500$) increase as the support $[a, 1]$ of the randomly generated efficiency becomes narrower, i.e., as the standard deviation of $\theta_j$ decreases.
This is expected since a lower standard deviation of the true efficiency makes it more difficult for the models to recover the true relative efficiency.

According to Table 5, the MHRS model also performs better than the HRS model. For low standard deviation of the true efficiency, i.e., for $\theta_j$ generated from $U[0.9, 1]$, the rank correlation in the MHRS model is at least 5% higher compared to the HRS model. However, for larger values of the standard deviation of $\theta_j$, the rank correlation in the MHRS model is only between 0.5% and 5% higher relative to the HRS model.

We obtain similar results for the case in which the theoretical efficiency $\theta_j$ is generated from the double truncated normal distribution $N_{[0,1]}(1, \sigma^2)$ defined on the interval $[0,1]$, with different values for $\sigma$. These results are presented in Appendix E.

To summarize, the Monte Carlo experiments show that, using the known DGP outlined above, the MHRS model consistently outperforms the conventional VRS and CRS models in recovering the true relative efficiency of DMUs. We also observe higher rank correlation by using the MHRS model compared to the HRS model, especially when the variance of the true efficiency is low.

7. Comparison with the Models of Cherchye et al. (2013)

Below we compare the MHRS model with the models of Cherchye et al. (2013). It is important to note that the specifications of these models are generally different. The models of Cherchye et al. (2013) are stated under the assumptions of either VRS or CRS, but do not allow a combination of both, as in the MHRS technology. However, the MHRS technology does not have a provision for allocating nonproportional inputs and outputs to individual processes as in the models of Cherchye et al. To enable comparison of the two approaches, below we consider the case in which there are no nonproportional inputs and outputs, i.e., the sets $I^{NP}$ and $O^{NP}$ are empty.

Furthermore, the approach of Cherchye et al. (2013) includes only joint inputs but not joint outputs. Incorporating the latter should be unproblematic but, for simplicity, below we assume that there are no shared outputs, i.e., that the set $O^S$ is empty.

Another important difference between the two approaches is in the interpretation and modeling of shared and joint inputs. The models of Cherchye et al. (2013) incorporate joint inputs under the
Table 6  DMU C generated by the assumption of joint input $x^S$.

<table>
<thead>
<tr>
<th>DMU</th>
<th>$x^1$</th>
<th>$x^2$</th>
<th>$x^S$</th>
<th>$y^1$</th>
<th>$y^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>$B$</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>$C$</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

assumption that they are simultaneously used by all processes and cannot be split between them in principle. In contrast, in the MHRS technology, it is assumed that shared inputs are \textit{allocated} to different processes. The MHRS technology is also suitable in a more general case in which inputs (and outputs) are partly shared and partly joint, or simply joint (see Endnote 3).

Example 3. The difference between the interpretation of joint and shared inputs leads to different models of technology. As an illustration, consider observed DMUs $A$ and $B$ in Table 6. Assume that there are two component processes, each characterized by the component-specific proportional input $x^k$ and output $y^k$, $k = 1, 2$. The third input $x^S$ is also proportional but not component-specific, and is used by both production processes.

According to the approach of Cherchye et al. (2013), if input $x^S$ is joint, i.e., simultaneously used by both processes, then DMU $C$ in Table 6 should be considered as producible. This DMU has the same process 1 as DMU $A$, and the same process 2 as DMU $B$, where both processes use the same joint input $x^S$. However, if input $x^S$ is shared, the above logic becomes unsubstantiated. Indeed, it is possible that DMU $A$ allocates the largest part of input $x^S$ to process 1, and DMU $B$ allocates the largest part of $x^S$ to process 2 (which allows them to produce high levels of outputs $y^1$ and $y^2$, respectively). In this case, assuming that DMU $C$ can produce both outputs $y^1$ and $y^2$ at the highest level would be incorrect, and this DMU is not included in the MHRS technology.

The above difference between the two approaches and its impact on efficiency evaluation is clarified further by a computational experiment discussed below. To compare the two approaches, we modify the DGP described in §6.1 by removing the shared output $y_j^S$ and the nonproportional input $x_j^{NP}$ and output $y_j^{NP}$. Because we no longer have the shared output, we correct the
DGP accordingly. Namely, the shared input (school expenditure) is now spent entirely on the two program-specific outputs \( y_1^j \) and \( y_2^j \). In this new DGP, the schools are generated as random vectors 
\[
(X_j, Y_j) = (x_1^j, x_2^j, x_S^j, y_1^j, y_2^j).
\]

The described DGP corresponds to the CRS model (LP-5) of Cherchye et al. (2013). We ran this model, and the MHRS model, in the output orientation. It is a simple mathematical fact that, in the described scenario, the CRS technology of Cherchye et al. (2013) is larger than the MHRS technology, for the reasons discussed in Example 3. Consequently, the efficiency of schools assessed in the former technology is never higher than in the latter.

However, because the MHRS technology is an inner approximation (a subset) of the true theoretical technology, the efficiency estimates \( \hat{\theta}_j \) of all schools calculated by the MHRS model are always conservative, i.e., we always have \( \hat{\theta}_j \geq \theta_j \), where \( \theta_j \) is the true efficiency of school \( j \) embedded in the DGP. In contrast, the CRS technology of Cherchye et al. (2013) is not a subset of the true technology from which the sample of schools is generated. In our calculations, for approximately 30% of all schools we had \( \tilde{\theta}_j < \theta_j \), where \( \tilde{\theta}_j \) is the output radial efficiency of school \( j \) in the CRS model of Cherchye et al. (2013). For all such schools, their output radial projections obtained from the latter model lie outside the assumed theoretical production technology. A numerical example in Appendix F provides an additional illustration of this last point.

Further computational results are presented in Appendix G. These results show that, for the assumed DGP described above (in which, importantly, \( x_S \) is a shared but not joint input), the MHRS model recovers the relative efficiency of DMUs better than the model of Cherchye et al. (2013). This advantage is more pronounced when the variation of the true efficiency \( \theta_j \) is relatively low. Thus, with \( \theta_j \) uniformly distributed in the interval [0.75, 1], the rank correlation between the true and estimated efficiency is approximately 10% higher when using the MHRS model compared to the model of Cherchye et al. (2013). With increasing standard deviation of the random efficiency \( \theta_j \) both models provide increasing average rank correlation. However, the advantage of the MHRS model over the model of Cherchye et al. (2013) is gradually reduced to about 1–3% in terms of rank correlation between the true and estimated efficiency.
In summary, as illustrated by Example 3, the assumption of joint inputs leads to a larger model of technology compared to the assumption that such inputs are shared (in the restricted setting when we have $I^N = O^N = O^S = \emptyset$). However, modeling shared inputs that are allocated to component processes (in unknown quantities) as joint inputs is theoretically unsubstantiated. As demonstrated by computations, in practice this leads to potential underestimation of the true efficiency of DMUs and obtaining their benchmarks outside the true technology.

8. Conclusion

In conventional CRS and VRS DEA models it is implicitly assumed that all inputs contribute to the production of all outputs. This generic nature of standard DEA models limits their usefulness in applications characterised by several component processes in which some inputs are used only for the production of specific outputs. For example, if some DMU $o$ has the largest ratio of some output $r$ to some input $i$ across all observed DMUs, then it is efficient in the CRS model, even if input $i$ is not used in the production of output $r$. This limitation becomes obvious in our motivational Example 1 involving the efficiency assessment of schools with two component processes. In this example, the use of the CRS model leads to a significant overestimation of school efficiency.

In our paper we address the above limitation and develop a new multiple hybrid returns-to-scale (MHRS) model of production technology with several component processes. This model is suitable for applications in which the inputs and outputs can be classed into the three types: proportional inputs and outputs specific to individual component processes, proportional inputs and outputs shared between all processes, and nonproportional inputs and outputs.

Our development is different from the existing approaches in several respects. First, we combine multiple proportional and nonproportional inputs and outputs in one model. We argue that this approach may, for example, be useful in the assessment of efficiency of schools, where both proportional (volume-related) and nonproportional (socio-economic and quality-related) measures occur naturally. Second, we derive the MHRS technology formally from explicitly stated production axioms. The resulting technology includes only those DMUs that can be explained using accepted
axioms, and does not include any arbitrary DMUs. Third, we do not make any explicit assumptions about the allocation of shared inputs and outputs to component production processes.

We illustrate the usefulness of the new MHRS technology by two large examples. The first is an application to public schools in Portugal providing secondary education. The second involves Monte Carlo experiments with a simulated DGP. In line with theoretical results and performance expectations, we show that the MHRS model provides better discrimination on efficiency and recovers the true theoretical efficiency of DMUs better than the alternative models.

It is worth noting that the MHRS technology is developed for a particular range of applications in which all Axioms 1–5 are satisfied. In particular, Axiom 5 requires that the individual component processes are scalable, which is similar to the assumption of CRS in a single-process technology. Clearly, in many applications, this assumption is not satisfied. In these cases the use of the MHRS model would be theoretically unsubstantiated.

The presented methodology suggests several avenues for future research. For example, a question arises if it is possible to develop an axiomatic definition of the MHRS technology that distinguishes between component-specific and shared nonproportional inputs and outputs. This approach should not require any information about the allocation of shared inputs and outputs, similar to the treatment of proportional inputs and outputs. It is also worthwhile exploring an axiomatic development of a technology in which different component processes are based on different assumptions of RTS. A further research avenue is the expansion of the MHRS model that would also include the type of joint inputs (and, potentially, outputs) in the sense of Cherchye et al. (2013).

Acknowledgments

The authors are grateful to the following colleagues and organisations for providing us with the data on Portuguese secondary schools: Luísa Loura, João Baptista and Joaquim Santos from Directorate-General for the Statistics of Education and Science (DGEEC), and José Passos and João Matos from Directorate-General for Planning and Financial Management (DGPGF), Ministry of Education and Science, Portugal.
Appendix A. Proofs of Proposition 1 and Theorem 1

Proof of Proposition 1. It suffices to consider the case \( 0 < \alpha_{\min} < 1 \). By Axiom 5 with \( \alpha = 0 \), the DMU \((X(0), Y(0))\), whose only nonzero component is vector \(X^{NP}\), is in \(T\). Define

\[
\left( \frac{\bar{X}}{\bar{Y}} \right) = \alpha_{\min} \left( \frac{X}{Y} \right) + (1 - \alpha_{\min}) \left( \frac{X(0)}{Y(0)} \right) = \left( \frac{\alpha_{\min} X^1, \ldots, \alpha_{\min} X^K, \alpha_{\min} X^S, X^{NP}}{\alpha_{\min} Y^1, \ldots, \alpha_{\min} Y^K, \alpha_{\min} Y^S, \alpha_{\min} Y^{NP}} \right).
\]

By Axiom 3, \((\bar{X}, \bar{Y}) \in T\). Define vector \(\tilde{\alpha} \in \mathbb{R}^K\) with components \(\tilde{\alpha}^k = \alpha^k / \alpha_{\min} \geq 1\), \(\forall k\). Then \(\tilde{\alpha}_{\min} = 1\) and \(\tilde{\alpha}_{\max} = \alpha_{\max} / \alpha_{\min}\). Applying Axiom 5 with \(\tilde{\alpha}\) to DMU \((\bar{X}, \bar{Y})\), we obtain \((\tilde{\lambda}(\bar{X}), \tilde{\lambda}(\bar{Y}))\). Hence the latter DMU is in \(T\). Conversely, \((\tilde{\lambda}(\alpha), \tilde{\lambda}(\alpha))\) dominates \((X(\alpha), Y(\alpha))\) in the Pareto sense. The proof follows by Axiom 2. \(\square\)

Proof of Theorem 1. Denote \(\tilde{T}^*\) the set of all DMUs \((X, Y) \geq 0\) that satisfy (4) with some vectors \(\lambda\) and \(\alpha_j, j \in J\). We need to prove that \(T^* = \tilde{T}^*\). The proof follows from Lemmas 1–3. \(\square\)

Lemma 1. Technology \(\tilde{T}^*\) satisfies Axioms 1–3.

Proof of Lemma 1. Any DMU \(j^* \in J\) satisfies (4) with \(\lambda_{j^*} = 1\), \(\lambda_j = 0\), \(\forall j \neq j^*, \alpha_j = 1\), \(\forall j, k\). Therefore, \(\tilde{T}^*\) satisfies Axiom 1. Axiom 2 is true because \((X', Y')\) satisfies (4) with the same vectors \(\lambda\) and \(\alpha_j, j \in J\), as the DMU \((X, Y)\). To prove Axiom 3, let \((\bar{X}, \bar{Y})\) satisfy (4) with vectors \(\hat{\lambda}\) and \(\hat{\alpha}_j, j \in J\). Also let \((\bar{X}, \bar{Y})\) satisfy (4) with \(\hat{\lambda}\) and \(\hat{\alpha}_j, j \in J\). Consider any \(\delta \in [0, 1]\). We need to prove that \((\bar{X}, \bar{Y}) = \delta(\bar{X}, \bar{Y}) + (1 - \delta)(\bar{X}, \bar{Y})\) satisfies (4) with some vectors \(\hat{\lambda}\) and \(\hat{\alpha}_j, j \in J\).

Let \(\bar{\lambda} = \delta \hat{\lambda} + (1 - \delta)\hat{\lambda}\). To define vectors \(\bar{\alpha}_j\), consider two cases. If \(\bar{\lambda}_j = 0\), we arbitrarily take \(\bar{\alpha}_j = 1\), \(\forall k\). If \(\bar{\lambda}_j > 0\), for each \(k\) define \(\bar{\alpha}_j^k\) so that \(\bar{\lambda}_j \bar{\alpha}_j^k = \delta \hat{\lambda}_j \alpha_j^k + (1 - \delta) \hat{\lambda}_j \bar{\alpha}_j^k\), i.e., \(\bar{\alpha}_j^k = \delta \hat{\lambda}_j \alpha_j^k + (1 - \delta) \hat{\lambda}_j \bar{\alpha}_j^k / \hat{\lambda}_j\). Let us prove that \((\bar{X}, \bar{Y})\) satisfies (4) with \(\bar{\lambda}\) and \(\bar{\alpha}_j, j \in J\). Conditions (4g) and (4h) are trivial. The proof of (4a) and (4d) is similar. For example, (4a) is proved as follows:

\[
\bar{X}^k = \delta \bar{X}^k + (1 - \delta) \bar{X}^k \geq \delta \sum_{j=1}^{n} \hat{\lambda}_j \max_k \{ \bar{\alpha}_j^k \} X_j^k + (1 - \delta) \sum_{j=1}^{n} \hat{\lambda}_j \max_k \{ \bar{\alpha}_j^k \} X_j^k = \sum_{j=1}^{n} \hat{\lambda}_j \max_k \{ \bar{\alpha}_j^k \} X_j^k.
\]

The proofs of (4b) and (4e) are also similar. Thus, for the former we have:

\[
\bar{X}^S \geq \delta \sum_{j=1}^{n} \hat{\lambda}_j \max_k \{ \bar{\alpha}_j^k \} X_j^S + (1 - \delta) \sum_{j=1}^{n} \hat{\lambda}_j \max_k \{ \bar{\alpha}_j^k \} X_j^S \geq \sum_{j=1}^{n} \left( \max_k \{ \delta \hat{\lambda}_j \alpha_j^k + (1 - \delta) \hat{\lambda}_j \bar{\alpha}_j^k \} \right) X_j^S = \sum_{j=1}^{n} \left( \max_k \{ \hat{\lambda}_j \alpha_j^k \} X_j^S \right) = \sum_{j=1}^{n} \hat{\lambda}_j \left( \max_k \{ \bar{\alpha}_j^k \} X_j^S \right).
\]
Finally, condition (4f) is proved as follows:

\[
\tilde{Y}^{NP} \leq \delta \sum_{j=1}^{n} \tilde{\lambda}_j \min \{1, \tilde{\alpha}_j^k \mid \forall k\} Y_j^{NP} + (1 - \delta) \sum_{j=1}^{n} \bar{\lambda}_j \min \{1, \alpha_j^k \mid \forall k\} Y_j^{NP}
\]

\[
\leq \sum_{j=1}^{n} \min \left\{ \delta \tilde{\lambda}_j + (1 - \delta) \bar{\lambda}_j, \delta \tilde{\lambda}_j \tilde{\alpha}_j^k + (1 - \delta) \bar{\lambda}_j \alpha_j^k \mid \forall k\right\} Y_j^{NP}
\]

\[
= \sum_{j=1}^{n} \min \{\tilde{\lambda}_j, \bar{\lambda}_j \tilde{\alpha}_j^k \mid \forall k\} Y_j^{NP} = \sum_{j=1}^{n} \bar{\lambda}_j (\min \{1, \tilde{\alpha}_j^k\} Y_j^{NP}).
\]

\[\Box\]

**Lemma 2.** Technology \(\tilde{T}^*\) satisfies Axiom 5.

**Proof of Lemma 2.** Let DMU \((X, Y)\) satisfy (4) with some \(\lambda\) and \(\alpha_j, j \in J\). We prove that DMU \((\tilde{X}(\beta), \tilde{Y}(\beta))\) defined in Axiom 5*, where \(\alpha\) is replaced by \(\beta = (\beta^1, \beta^2, \ldots, \beta^K)\), satisfies (4) with the same vector \(\lambda\) and vectors \(\tilde{\alpha}_j = (\beta^1 \alpha_j^1, \beta^2 \alpha_j^2, \ldots, \beta^K \alpha_j^K), j \in J\). The proof of conditions (4a), (4c), (4d), (4g) and (4h) is straightforward. For example, in the case of (4a), for each \(k\) we have:

\[
\tilde{X}^k(\beta) = \beta^k X^k \geq \beta^k \sum_{j=1}^{n} \lambda_j (\alpha_j^k X_j^k) = \sum_{j=1}^{n} \lambda_j (\tilde{\alpha}_j^k X_j^k).
\]

Condition (4b) follows from the following inequality:

\[
\tilde{X}^S(\beta) = \max_k \{\beta^k\} X^S \geq \left(\max_k \{\beta^k\}\right) \sum_{j=1}^{n} \lambda_j \left(\max_k \{\alpha_j^k\} X_j^S\right)
\]

\[
\geq \sum_{j=1}^{n} \lambda_j \left(\max_k \{\beta^k \alpha_j^k\} X_j^S\right) = \sum_{j=1}^{n} \lambda_j \left(\max_k \{\tilde{\alpha}_j^k\} X_j^S\right).
\]

The proof of inequality (4e) is similar. Finally, (4f) is proved as follows:

\[
\tilde{Y}^{NP}(\beta) = \min \{1, \beta^k \mid \forall k\} Y^{NP} \leq \min \{1, \beta^k \mid \forall k\} \sum_{j=1}^{n} \lambda_j (\min \{1, \alpha_j^k \mid \forall k\} Y_j^{NP})
\]

\[
\leq \sum_{j=1}^{n} \lambda_j (\min \{1, \beta^k \alpha_j^k \mid \forall k\} Y_j^{NP}) = \sum_{j=1}^{n} \lambda_j (\min \{1, \tilde{\alpha}_j^k \mid \forall k\} Y_j^{NP}).
\]

\[\Box\]

**Lemma 3.** \(\tilde{T}^* \subseteq T\), where \(T\) is any technology that satisfies Axioms 1–3 and 5.

**Proof of Lemma 3.** Let \((X, Y) \in \tilde{T}^*\). We need to prove that \((X, Y) \in T\). By definition of \(\tilde{T}^*\), \((X, Y)\) satisfies (4) with some \(\lambda\) and \(\alpha_j, j \in J\). The left-hand side of conditions (4a)–(4f) is a convex
Appendix B. Proofs of Theorem 2 and Propositions 2–4

In the proofs below we denote \( \tilde{T}_{\text{MHRS}} \) the set of all DMUs \( (X,Y) \geq 0 \) that satisfy conditions (6) with some vectors \( \lambda \) and \( \mu_j, j \in J \). We first establish some preliminary results.

**Lemma 4.** \( T^* \subseteq \tilde{T}_{\text{MHRS}} \).

**Proof of Lemma 4.** Let \( (X,Y) \in T^* \). Then \( (X,Y) \) satisfies (4) with some vectors \( \lambda \) and \( \alpha_j, j \in J \). Then \( (X,Y) \) satisfies (6) with the same vector \( \lambda \) and \( \mu_j^k = \lambda_j \alpha_j^k - \lambda_j \), as in (5). Indeed, substituting \( \lambda_j \alpha_j^k \) by \( \lambda_j + \mu_j^k \) in (4), we obtain (6). In particular, (6f) is true because

\[
\lambda_j (\min\{1, \alpha_j^m\}) = \min \{\lambda_j, \lambda_j \alpha_j^k \mid \forall k\} = \min \{\lambda_j, \lambda_j + \mu_j^k \mid \forall k\} = \lambda_j + \min \{0, \mu_j^k\}, \quad \forall j. \tag{12}
\]

We have proved that \( (X,Y) \in \tilde{T}_{\text{MHRS}} \). \( \square \)

**Lemma 5.** \( \tilde{T}_{\text{MHRS}} \subseteq \text{cl} T^* \). In other words, if \( (X,Y) \in \tilde{T}_{\text{MHRS}} \), then either \( (X,Y) \in T^* \) or \( (X,Y) \) is a limit point of \( T^* \).

**Proof of Lemma 5.** Let \( (X,Y) \) satisfy (6) with some vectors \( \lambda \) and \( \mu_j, j \in J \). Two cases arise.

1) Assume that, if \( \lambda_j = 0 \) then \( \mu_j^k = 0, \forall j, k \). For each \( j \) such that \( \lambda_j > 0 \) we use (5) to define \( \alpha_j^k = (\lambda_j + \mu_j^k) / \lambda_j^k, \forall k \). If \( \lambda_j = 0 \), arbitrarily define \( \alpha_j^k = 1, \forall k \). Then (5) is true for all \( j \) and \( k \), and conditions (6) imply (4). In particular, (4f) follows from (12). Therefore, \( (X,Y) \in T^* \).

2) Assume \( \lambda_j = 0 \) and \( \mu_j^k > 0 \) for some \( j \) and \( k \). (By (6h), the case \( \mu_j^k < 0 \) is impossible.) Let us show that \( (X,Y) \) is the limit of some sequence of DMUs \( (X^t,Y^t) \in T^*, t = 1,2, \ldots \) Let \( (\tilde{X}, \tilde{Y}) = (1/n) \sum_{j \in J}(X_j, Y_j) \) be the average of all observed DMUs. For each \( t \), define the convex combination

\[
(X^t,Y^t) = (1 - 1/t)(X,Y) + (1/t)(\tilde{X}, \tilde{Y}) \overset{t \to +\infty}{\longrightarrow} (X,Y).
\]

We need to prove that \( (X^t,Y^t) \in T^*, \forall t \). Each observed DMU \( (X_{j^*}, Y_{j^*}) \), \( j^* \in J \), satisfies (6) with \( \lambda_{j^*} = 1, \lambda_j = 0, \forall j \neq j^* \), and vectors \( \mu_j = 0, \forall j \). As assumed, \( (X,Y) \) satisfies (6) with vectors
λ and μ_j, ∀j. Then, for each t, (X^t, Y^t) satisfies (6) with the vectors \( \lambda_j^{(t)} = (1 - 1/t)\lambda_j + 1/(tn) \) and \( \mu_j^{(t)} = (1 - 1/t)\mu_j, \forall j \in J \). Therefore, (X^t, Y^t) ∈ \( \tilde{T}_{\text{MHRS}} \), ∀t. As proved in the above case 1), because \( \lambda_j^{(t)} > 0, \forall j \), we have (X^t, Y^t) ∈ \( T^* \), ∀t. Therefore, (X, Y) is a limit point of \( T^* \).

**Lemma 6.** A DMU \((X, Y)\) satisfies conditions (6) with some vectors \( \lambda, \mu_j, j \in J \), if and only if \((X, Y)\) satisfies conditions (7) with some vectors \( \lambda, \rho, \sigma, \tau \) and \( \mu_j, j \in J \).

**Proof of Lemma 6.** Let \((X, Y)\) satisfy (6) with some vectors \( \lambda \) and \( \mu_j, j \in J \). For all \( j \in J \), define

\[
\rho_j = \max_k \{ \mu_j^k \mid k \in K \}, \quad \sigma_j = \min_k \{ \mu_j^k \mid k \in K \}, \quad \tau_j = -\min \{ 0, \mu_j^k \mid k \in K \}.
\]

Then \((X, Y)\) satisfies (7) with \( \lambda, \rho, \sigma, \tau \) and \( \mu_j, j \in J \). Conversely, let \((X, Y)\) satisfy (7) with some \( \lambda, \rho, \sigma, \tau \) and \( \mu_j, j \in J \). By (7h), \( \lambda \geq 0 \). Also, (7b) and (7j) imply (6b). Similarly, (7e) and (7k) imply (6e), (7f) and (7i) imply (6f). Finally, (7h) and (7i) imply (6h).

**Lemma 7.** Technology \( \tilde{T}_{\text{MHRS}} \) is a polyhedral set. If \( I^{\text{NP}} = O^{\text{NP}} = \emptyset \), \( \tilde{T}_{\text{MHRS}} \) is a polyhedral cone.

**Proof of Lemma 7.** By Lemma 6, let \( \tilde{T}_{\text{MHRS}} \) be defined by conditions (7). Denote \( W \) the set of all solutions \( (X, Y, \lambda, \rho, \sigma, \tau, \mu_j \mid \forall j) \) that satisfy conditions (7) and the nonnegativity conditions \( X, Y \geq 0 \). The set \( W \) is defined by a finite number of linear inequalities and is, therefore, a polyhedral set. Technology \( \tilde{T}_{\text{MHRS}} \) is the projection of \( W \) on its dimensions \( X \) and \( Y \). By the projection lemma (see, e.g., its use for the proof of a similar Proposition 1 in Podinovski et al. 2016), \( \tilde{T}_{\text{MHRS}} \) is a polyhedral set.

Suppose that \( I^{\text{NP}} = O^{\text{NP}} = \emptyset \). Let DMU \((X, Y)\) satisfy (6) from which the constraints (6c) and (6f) are omitted, with some vectors \( \tilde{\lambda} \) and \( \tilde{\mu}_j, j \in J \). Then, for any scalar \( \alpha \geq 0 \), \((\alpha X, \alpha Y)\) satisfies (6) with \( \hat{\lambda} = \tilde{\lambda} \) and \( \hat{\mu}_j = \alpha \mu_j + (\alpha - 1)\tilde{\lambda} \), for which \( \hat{\lambda} + \mu_j = \alpha (\tilde{\lambda} + \mu_j), \forall j \in J \). Therefore \( T_{\text{MHRS}} \) is a cone.

**Lemma 8.** \( \tilde{T}_{\text{MHRS}} = \text{cl} \ T^* \).

**Proof of Lemma 8.** By Lemma 4, \( T^* \subseteq \tilde{T}_{\text{MHRS}} \). By Lemma 7, \( \tilde{T}_{\text{MHRS}} \) is a closed set. Therefore, \( \text{cl} \ T^* \subseteq \tilde{T}_{\text{MHRS}} \). Taking into account Lemma 5, \( \tilde{T}_{\text{MHRS}} = \text{cl} \ T^* \).
Lemma 9. Technology $\tilde{T}_{\text{MHRS}}$ satisfies Axioms 1–5.

Proof of Lemma 9. Axioms 1 and 2 are trivial. By Lemma 7, $\tilde{T}_{\text{MHRS}}$ is a polyhedral set. Therefore it satisfies Axioms 3 and 4. To prove Axiom 5, let $(X,Y) \in \tilde{T}_{\text{MHRS}}$ satisfy (6) with some $\lambda$ and $\mu_j$, $j \in J$. Consider any $\alpha \in \mathbb{R}_+^k$. We need to prove that $(X(\alpha),Y(\alpha))$ in (3) satisfies (6) with some $\tilde{\lambda}$ and $\tilde{\mu}_j$, $j \in J$. Indeed, let $\tilde{\lambda} = \lambda$. Define $\tilde{\mu}_j^k = \alpha^k(\lambda_j + \mu_j^k) - \lambda_j$ so that $\tilde{\lambda}_j + \tilde{\mu}_j^k = \alpha^k(\lambda_j + \mu_j^k)$, $\forall j, k$. Let us prove that $(X(\alpha),Y(\alpha))$ satisfies (6) with $\tilde{\lambda}$ and $\tilde{\mu}_j$, $j \in J$. Conditions (6g)–(6i) are trivial. The proof of (6a) and (6d) is similar. For example, in the case of (6a) we have:

$$X^k(\alpha) = \alpha^k X^k \geq \sum_{j=1}^n \alpha^k (\lambda_j + \mu_j^k) X_j^k = \sum_{j=1}^n (\tilde{\lambda}_j + \tilde{\mu}_j^k) X_j^k.$$

The proof of (6b) and (6e) is similar. For example, the former is proved as follows:

$$X^S(\alpha) = \left(\max_k \{\alpha^k\}\right) X^S \geq \left(\max_k \{\alpha^k\}\right) \sum_{j=1}^n \max_k \{\lambda_j + \mu_j^k\} X_j^S \geq \sum_{j=1}^n \max_k \{\alpha^k (\lambda_j + \mu_j^k)\} X_j^S = \sum_{j=1}^n \max_k \{\tilde{\lambda}_j + \tilde{\mu}_j^k\} X_j^S.$$

Condition (6c) is true because $\tilde{\lambda} = \lambda$ and $X(\alpha)^{NP} = X^{NP}$. Finally, to prove (6f), consider two cases. If $\alpha^{\text{min}} < 1$ then $Y^{NP}(\alpha) = 0$ and (6f) whose left-hand side is nonnegative follows. If $\alpha^{\text{min}} \geq 1$ then $Y^{NP}(\alpha) = Y^{NP}$. Then (6f) follows from (6h) and the definition of $\tilde{\lambda}$ and $\tilde{\mu}_j^k$. □

Lemma 10. $\tilde{T}_{\text{MHRS}} \subseteq T$, where $T$ is any technology that satisfies Axioms 1–5.

Proof of Lemma 10. Because $T$ satisfies Axioms 1–3 and 5, by Definition 1, $T^* \subset T$. By Axiom 4, $T$ is closed and thus $\text{cl } T^* \subset T$. By Lemma 5, $\tilde{T}_{\text{MHRS}} \subseteq \text{cl } T^* \subset T$. □

Proof of Theorem 2. This follows from Lemmas 9 and 10. □

Proof of Propositions 2, 3 and 4. By Theorem 2, $T_{\text{MHRS}} = \tilde{T}_{\text{MHRS}}$. The rest of the proof follows from Lemmas 6, 7 and 8, respectively. □

Appendix C. Proof of Theorem 3

Proof of Theorem 3. Consider the dual to program (8) in which $\lambda$ is sign free. (Condition $\lambda \geq 0$ follows from (7h) and is redundant.) Let $v \in \mathbb{R}_+^m$ and $u \in \mathbb{R}_+^s$ be the dual vectors to the input and output constraints in (8), respectively. Let $\omega$ be dual to (7g) and $\delta \in \mathbb{R}_+^n$ be dual to (7h). Denote
ε^k_j, ξ^k_j, ζ^k_j ≥ 0 the dual variables dual to (7i), (7j) and (7k), ∀j,k. The dual to (8) is stated as follows. (Its constraints correspond to following order of primal variables: η, λ^j_j, μ^k_j, τ^j_j, ρ^j_j and σ^j_j.)

\[
\begin{align*}
\min & \quad v^\top X_o + \omega & (13a) \\
\text{s.t.} & \quad u^\top Y_o = 1, & (13b) \\
& \quad v^\top X_j - u^\top Y_j + \omega - \delta_j = 0, & \forall j, (13c) \\
& \quad \sum_{i\in I^k} v_i X_{ji} - \sum_{r\in O^k} u_r Y_{jr} - \varepsilon^k_j + \xi^k_j - \zeta^k_j = 0, & \forall j,k, (13d) \\
& \quad \sum_{r\in O^{NP}} u_r Y_{jr} + \delta_j - \sum_{k\in K} \varepsilon^k_j ≥ 0, & \forall j, (13e) \\
& \quad \sum_{i\in I^S} v_i X_{ji} - \sum_{k\in K} \xi^k_j = 0, & \forall j, (13f) \\
& \quad - \sum_{r\in O^{S}} u_r Y_{jr} + \sum_{k\in K} \zeta^k_j = 0, & \forall j, (13g) \\
& \quad u, v ≥ 0, \delta_j, \varepsilon^k_j, \xi^k_j, \zeta^k_j ≥ 0, \forall j,k, \omega \text{ sign free.} & (13h)
\end{align*}
\]

Equalities (13f) and (13g) are equivalently restated as nonlinear constraints:

\[
\begin{align*}
\xi^k_j &= p^k_j \sum_{i\in I^S} v_i X_{ji}, \text{ where } \sum_{k\in K} p^k_j = 1, p^k_j ≥ 0, \forall j,k, \\
\zeta^k_j &= q^k_j \sum_{r\in O^{S}} u_r Y_{jr}, \text{ where } \sum_{k\in K} q^k_j = 1, q^k_j ≥ 0, \forall j,k.
\end{align*}
\] (14)

Substitute \(\xi^k_j\) and \(\zeta^k_j\) from (14) into (13d) and \(\delta_j\) from (13c) into (13e). Rearranging, we have

\[
\begin{align*}
\min & \quad v^\top X_o + \omega \\
\text{s.t.} & \quad u^\top Y_o = 1, \\
& \quad v^\top X_j - u^\top Y_j + \omega = \delta_j ≥ 0, & \forall j, \\
& \quad \left(\sum_{i\in I^k} v_i X_{ji} + p^k_j \sum_{i\in I^S} v_i X_{ji}\right) - \left(\sum_{r\in O^k} u_r Y_{jr} + q^k_j \sum_{r\in O^{S}} u_r Y_{jr}\right) = \varepsilon^k_j ≥ 0, & \forall j,k \\
& \quad \sum_{i\in I^S} v_i X_{ji} - \sum_{r\in O^{NP}} u_r Y_{jr} + \omega ≥ 0, & \forall j, \\
& \quad \sum_{k\in K} p^k_j = \sum_{k\in K} q^k_j = 1, & \forall j, \\
& \quad u, v ≥ 0, p^k_j, q^k_j ≥ 0, \forall j,k, \omega \text{ sign free.}
\end{align*}
\]

Removing variables \(\delta_j\) and \(\varepsilon^k_j\) as redundant, we obtain program (9). \(\square\)
Appendix D. The Data Generating Process for Section 6

We generate the vector of teaching hours for the two programs in polar coordinates, i.e., we let 
\[(x_j^1, x_j^2) = (r_j \cos \varphi_j, r_j \sin \varphi_j),\]
where the angle \(\varphi_j\) and radius \(r_j\) are randomly generated from the uniform distributions \(U[\frac{\pi}{4} - 0.3, \frac{\pi}{4} + 0.3]\) and \(U[1,100]\), respectively. Similarly, we generate the vector of expenditures for the two programs as 
\[(x_j^S1, x_j^S2) = (\rho_j \cos \psi_j, \rho_j \sin \psi_j),\]
where \(\psi_j\) and \(\rho_j\) are generated from the distributions \(U[\frac{\pi}{4} - 0.3, \frac{\pi}{4} + 0.3]\) and \(U[0.005,5]\), respectively.

Next, we generate the proportion \(\pi_j^0 \sim U[0,0.1]\) of the teaching hours, and the proportion \(\pi_j^S \sim U[0,0.1]\) of the expenditure allocated to the students with special needs, in addition to the regular teaching hours and expenditure. These proportions are assumed to be the same for both programs 1 and 2 of school \(j\). The proportions \(1 - \pi_j^0\) and \(1 - \pi_j^S\) represent the regular teaching hours and expenditure for all students on both programs. We now calculate the number of all students \(y_j^k\) and students with special needs \(y_j^{Sk}\) on each program \(k = 1, 2\) using the following functions:

\[
y_j^k = ((1 - \pi_j^0)x_j^k \times (1 - \pi_j^S)x_j^{Sk})^{0.5}, \quad k = 1, 2,
\]
\[
y_j^{Sk} = 0.5((\pi_j^0)x_j^k \times \pi_j^Sx_j^{Sk})^{0.5}, \quad k = 1, 2.
\]

We generate the random output radial efficiency \(\theta_j \in [0,1]\) and define \(\tilde{\theta}_j^k = \theta_jy_j^k, \tilde{\theta}_j^{Sk} = \theta_jy_j^{Sk}\), \(k = 1, 2\). For different experiments, we generate \(\theta_j\) from the uniform distribution \(U[a,1]\), where \(a \in \{0.25,0.5,0.75,0.9\}\), and from the normal distribution \(N[0.1](1,\sigma^2)\) truncated to the range \([0,1]\), where \(\sigma \in \{0.04,0.12,0.24,0.4\}\).

Next, we randomly generate the proportion \(\tilde{\pi}_j \sim U[0.01,0.09]\) of all students assumed to have good exam results on entry to both programs. The number of such students is calculated as

\[x_j^{NP} = \tilde{\pi}_j\theta_j(y_j^1 + y_j^2).\]

For a school operating efficiently, the total number of students with good results on exit is then assumed to be as follows:

\[y_j^{NP} = (x_j^{NP})^{0.8}(x_j^1 + x_j^2)^{0.1}.\]

Applying efficiency \(\theta_j\), we obtain \(\tilde{y}_j^{NP} = \theta_jy_j^{NP}\). We also define the total school expenditure as \(x_j^S = x_j^{S1} + x_j^{S2}\) and the total number of students with special needs as \(\tilde{y}_j^S = \tilde{y}_j^{S1} + \tilde{y}_j^{S2}\). The final inputs and outputs for the generated school \(j\) that are used in the simulation analysis are as follows:

\[(x_j^1, x_j^2, x_j^S, x_j^{NP}, y_j^1, y_j^2, y_j^S, \tilde{y}_j^{NP}).\]
It is straightforward to show that, given the specification of production functions (15) and (16), the theoretical technology $T_0$ of which the schools (17) are random members, satisfies all Axioms 1–5. In particular, technology $T_0$ is convex and exhibits multiple selective proportionality as specified by Axiom 5. This implies that the MHRS model of technology $T_0$ based on the final sample of generated schools (and, consequently, the VRS and HRS models that are defined by weaker sets of axioms) is an inner approximation (i.e., a subset) of this technology. However, the standard CRS model is not an inner approximation of $T_0$, because $T_0$ is not a cone technology.

**Appendix E. Computations with the Double Truncated Normal Distribution**

Table 7 provides an additional illustration for §6.2. It shows the computational results obtained for the efficiency $\theta_j$ randomly generated from the double truncated normal distribution $N_{[0,1]}(1, \sigma^2)$ truncated to $[0,1]$, where $\sigma \in \{0.04, 0.12, 0.24, 0.4\}$, and for sample sizes $n \in \{50, 100, 200, 500\}$.

**Table 7** The average rank correlation coefficients (taken over 100 replications) between the true and estimated efficiency scores in different models. The true theoretical efficiency is generated from the double truncated normal distribution. The standard deviations of the rank correlation coefficients are shown in parentheses.
The interpretation of results in Table 7 is similar to the case of uniform distribution of $\theta_j$ discussed in §6.2. As in the latter case, we observe that the MHRS model recovers the true relative efficiency (ranks) better than the VRS, CRS and HRS models.

**Appendix F. Example for Section 7**

Consider technology $T_1$ with two component processes $k = 1, 2$. Process 1 uses two inputs (e.g., labor and costs) denoted $x^1$ and $x^{S1}$. Similarly, process 2 uses inputs $x^2$ and $x^{S2}$. The two processes produce component-specific outputs $y^k$, $k = 1, 2$. Define technology $T_1$ as follows:

$$T_1 = \left\{ (x^1, x^2, x^{S1}, x^{S2}, y^1, y^2) \in \mathbb{R}^4_+ \times \mathbb{R}^2_+ \mid y^1 \leq (x^1 x^{S1})^{0.5}, y^2 \leq (x^2 x^{S2})^{0.5} \right\}.$$

Note that technology $T_1$ is a convex cone. The following three DMUs are clearly efficient in technology $T_1$: $A = (1, 1, 1, 1, 1, 1)$, $B = (1, 0, 2, 0, \sqrt{2}, 0)$ and $C = (0, 1, 0, 2, 0, \sqrt{2})$.

We now add the inputs $x^{S1}$ and $x^{S2}$ thus creating a shared input $x^S = x^{S1} + x^{S2}$. This transforms technology $T_1$ to technology $T_2$. In the latter, DMUs are stated in the form $(x^1, x^2, x^S, y^1, y^2)$. This transforms DMUs $A$, $B$ and $C$ to $\tilde{A} = (1, 1, 2, 1, 1)$, $\tilde{B} = (1, 0, 2, \sqrt{2}, 0)$ and $\tilde{C} = (0, 1, 2, 0, \sqrt{2})$.

It is straightforward to verify that technology $T_2$ is convex and satisfies Axiom 5 if we treat $x^1$, $x^2$, $y^1$ and $y^2$ as component-specific proportional inputs and outputs, respectively, and $x^S$ as a shared proportional input. Furthermore, DMUs $\tilde{A}$, $\tilde{B}$ and $\tilde{C}$ are efficient in technology $T_2$.

Consider the MHRS technology $T_{\text{MHRS}}$ generated by DMUs $\tilde{A}$, $\tilde{B}$ and $\tilde{C}$. Because technology $T_2$ satisfies Axioms 1–5, and because $T_{\text{MHRS}}$ is the smallest technology that satisfies them, we have $T_{\text{MHRS}} \subseteq T_2$. Therefore, DMUs $\tilde{A}$, $\tilde{B}$ and $\tilde{C}$ are efficient in technology $T_{\text{MHRS}}$.

Consider the output-oriented analogue of model (LP-5) in Cherchye et al. (2013). DMU $\tilde{A}$, which is theoretically efficient, is inefficient in this model, and its output radial efficiency is equal to $1/\sqrt{2}$. Indeed, in this model the input $x^S = 2$ is treated as the joint input that is used as a whole (without splitting) by both processes 1 and 2. Under this assumption, processes 1 and 2 of DMU $\tilde{A}$ are outperformed by the single DMUs $\tilde{B}$ and $\tilde{C}$, respectively. This implies that the output radial target for DMU $\tilde{A}$ is DMU $A^* = (1, 1, 2, \sqrt{2}, \sqrt{2})$, which is located outside the technology $T_2$. 

Appendix G. Comparison with the Model of Cherchye et al. (2013)

Below we provide additional computational results that support our discussion in §7. The Monte Carlo study used to compare the MHRS model with the model (LP-5) of Cherchye et al. (2013) is similar to the study reported in Table 5, but is based on the simplified DGP described in §7.

Table 8 The average rank correlation coefficients (taken over 100 replications) between the true and estimated efficiency scores in different models. The standard deviations of the rank correlation coefficients are shown in parentheses.

<table>
<thead>
<tr>
<th>Efficiency</th>
<th>n</th>
<th>Cherchye et al.</th>
<th>MHRS</th>
<th>CRS</th>
</tr>
</thead>
<tbody>
<tr>
<td>U[0.9,1]</td>
<td>50</td>
<td>0.4434 (0.1332)</td>
<td>0.636 (0.09513)</td>
<td>0.4052 (0.1113)</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.4924 (0.08341)</td>
<td>0.7271 (0.06286)</td>
<td>0.4883 (0.07223)</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.5197 (0.05882)</td>
<td>0.7988 (0.03851)</td>
<td>0.5774 (0.05241)</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.5524 (0.03149)</td>
<td>0.874 (0.01918)</td>
<td>0.6883 (0.02895)</td>
</tr>
<tr>
<td>StDev ≈ 0.03</td>
<td>100</td>
<td>0.7366 (0.0823)</td>
<td>0.8072 (0.0688)</td>
<td>0.6104 (0.09256)</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.8252 (0.03316)</td>
<td>0.909 (0.02394)</td>
<td>0.7501 (0.04443)</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.8594 (0.01693)</td>
<td>0.9458 (0.01149)</td>
<td>0.8271 (0.02157)</td>
</tr>
<tr>
<td>U[0.75,1]</td>
<td>50</td>
<td>0.8709 (0.0547)</td>
<td>0.8928 (0.04874)</td>
<td>0.7294 (0.08093)</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.9062 (0.03815)</td>
<td>0.93 (0.03397)</td>
<td>0.7867 (0.05986)</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.9335 (0.02019)</td>
<td>0.956 (0.01673)</td>
<td>0.8364 (0.03795)</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.9539 (0.00992)</td>
<td>0.9753 (0.007664)</td>
<td>0.8841 (0.01685)</td>
</tr>
<tr>
<td>StDev ≈ 0.07</td>
<td>100</td>
<td>0.7878 (0.0546)</td>
<td>0.8687 (0.04562)</td>
<td>0.6803 (0.0587)</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.8252 (0.03316)</td>
<td>0.909 (0.02394)</td>
<td>0.7501 (0.04443)</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.8594 (0.01693)</td>
<td>0.9458 (0.01149)</td>
<td>0.8271 (0.02157)</td>
</tr>
<tr>
<td>U[0.5,1]</td>
<td>50</td>
<td>0.8709 (0.0547)</td>
<td>0.8928 (0.04874)</td>
<td>0.7294 (0.08093)</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.9062 (0.03815)</td>
<td>0.93 (0.03397)</td>
<td>0.7867 (0.05986)</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.9335 (0.02019)</td>
<td>0.956 (0.01673)</td>
<td>0.8364 (0.03795)</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.9539 (0.00992)</td>
<td>0.9753 (0.007664)</td>
<td>0.8841 (0.01685)</td>
</tr>
<tr>
<td>StDev ≈ 0.14</td>
<td>100</td>
<td>0.9062 (0.03815)</td>
<td>0.93 (0.03397)</td>
<td>0.7867 (0.05986)</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.9335 (0.02019)</td>
<td>0.956 (0.01673)</td>
<td>0.8364 (0.03795)</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.9539 (0.00992)</td>
<td>0.9753 (0.007664)</td>
<td>0.8841 (0.01685)</td>
</tr>
</tbody>
</table>

Table 8 reports the results from the Monte Carlo analysis in which the true efficiency $\theta_j$ is generated from the uniform distribution $U[a, 1]$ with the lower bound $a \in \{0.25, 0.5, 0.75, 0.9\}$, and for sample sizes $n \in \{50, 100, 200, 500\}$. For each combination of sample size $n$ and lower bound $a$, we estimate the efficiency of the $n$ DMUs using the MHRS model, the model (LP-5) of Cherchye et al. (2013), and, for reference purposes, the conventional CRS model, all in the output orientation.

We again use rank correlation as a measure of the model’s ability to recover the true relative efficiency scores. Table 8 reports the average rank correlation between the estimated and the true
efficiency scores based on 100 replications. The structure of this table is identical to Table 5. As expected, we see that the rank correlation increases for all three models as the sample size $n$ increases.

These results clearly show that the MHRS model outperforms both the model of Cherchye et al. (2013) and the CRS model. This is especially pronounced when the standard deviation of the random efficiency is relatively low. With the standard deviation below 0.07, the rank correlation of the MHRS model is at least 10% higher compared to the model of Cherchye et al. (2013). This advantage diminishes to about 1–3% when the variance of the true efficiency is increased, i.e., when $\theta_j$ is generated from the uniform distribution $U[a, 1]$, where $a = 0.5$ or $a = 0.25$. The standard deviations of the rank correlation in Table 8 are relatively low which indicates a relatively low dispersion around the average rank correlation.
Endnotes

1. It is known that, if DMU\(_o\) has the largest ratio of some output \(r\) to some input \(i\) among all observed DMUs, then DMU\(_o\) is efficient in the CRS model, even if input \(i\) is not used in the production of output \(r\). For example, school \(A\) is efficient because its ratio of the language students to mathematics teachers is higher than the same ratio for school \(B\).

2. In this paper we assume that all inputs and outputs are given as volume measures and not as percentages or ratios. An example of a volume input is the number of students with good grades on entry, and an example of a ratio input is the percentage of such students in the entire intake. A principal problem with ratio measures is that they generally do not satisfy the standard convexity assumption and need to be modeled differently to the volume measures. A range of DEA models with a single proportional process and nonproportional ratio inputs and outputs is developed by Olesen et al. (2015, 2017) who extend earlier ideas of Ruggiero (1996) and Podinovski (2005).

3. The worst-case assumption underpinning the development of the MHRS technology is applicable to a range of different types of shared inputs. On the two sides of this range are the fully allocated (although in unknown proportions) and perfectly joint inputs (the latter in the sense of Cherchye et al. 2013). This range also includes any combination of the two extreme types, which often occurs in practical applications because of insufficient information (see our discussion of school costs in §1.1). As shown by Example 3 in §7, if all shared inputs are joint, the assumption that such inputs be scaled by the maximum of the scaling factors of individual processes is over-conservative (but it is still valid and can be used, in the same way as the VRS technology can be used to model a CRS production process). In this case, the approach of Cherchye et al. (2013) utilizes the information that the inputs are actually joint and provides a more appropriate treatment. However, the approach of Cherchye et al. (2013) is not applicable if some shared inputs are fully or partly allocated to different processes.

4. In the axiomatic definition of the VRS and CRS technology (Banker et al. 1984), Axiom 4 follows from the other accepted axioms and for this reason does not need to be stated. However, as shown in §3, the technology \(T^*\) defined by Axioms 1–3 and 5 only is not closed.
5. It is clear that not knowing the exact proportions in which the shared inputs and outputs are allocated to component processes and having to use the worst-case assumption as a substitute, represents a loss of information leading to a weaker (smaller) model of technology compared to the case in which the actual allocations are known. For example, consider the technology with two component processes and a single shared input. Let DMU \((x^1, x^2, x^S, y^1, y^2) = (1, 1, 1, 1, 1)\). Suppose that the shared input \(x^S = 1\) is actually split equally between the two processes. Then, scaling the process 1 and its share of \(x^S\) by a factor \(\alpha^1 = 2\), and keeping process 2 unchanged, we obtain DMU \(A = (2, 1, 1.5, 2, 1)\). However, by the worst-case assumption implemented in Axiom 5, we obtain DMU \(B = (2, 1, 2, 2, 1)\). Because \(A\) outperforms \(B\), and because we assume Axiom 2, the former approach leads to a larger technology than the latter.

6. To see that technology \(T^*\) is generally not closed, consider the following example with \(K = 1\). Let \(I^S = O^S = O^{NP} = \emptyset\). Define two observed DMUs as \((X_1, Y_1) = (1, 2, 2)\) and \((X_2, Y_2) = (1, 1, 1)\), where the three vector components represent \(i_1 \in I^1, i_2 \in I^{NP}\) and \(r_1 \in O^1\), respectively. By Axiom 5, the DMUs \((X_1(t), Y_1(t)) = (t, 2, 2t) \in T^*\), for all \(t = 1, 2, \ldots\). By Axiom 3, the convex combination

\[
\begin{pmatrix} X' \\ Y' \end{pmatrix} = \frac{1}{t} \begin{pmatrix} X_1(t) \\ Y_1(t) \end{pmatrix} + \left(1 - \frac{1}{t}\right) \begin{pmatrix} X_2 \\ Y_2 \end{pmatrix} = \begin{pmatrix} (2t - 1)/t \\ (t + 1)/t \end{pmatrix} \in T^*, \ \forall t.
\]

Note that \(\{(X', Y')\} \quad \xrightarrow{t \to +\infty} (2, 1, 3)\). Let us show that \((2, 1, 3) \notin T^*\). Assume this is not true. Then \((2, 1, 3)\) satisfies (4) with some \(\lambda_1, \lambda_2, \alpha_1^1\) and \(\alpha_2^1\). From (4c) we have \(2\lambda_1 + 1\lambda_2 \leq 1\). Taking into account equality (4g), this implies \(\lambda_1 = 0\) and \(\lambda_2 = 1\). Then (4a) and (4d) imply \(\alpha_2^1 \leq 2\) and \(\alpha_2^1 \geq 3\), respectively, which is impossible.

7. If \(I^{NP} = O^{NP} = \emptyset\), the constraints (6c) and (6f) are omitted from the statement (6) of technology \(T_{MHRS}\), and the remaining conditions are simplified by the substitution \(\hat{\lambda}_j^k = \lambda_j + \mu_j^k\), \(\forall j, k\). Furthermore, the normalizing condition (6g) becomes redundant and is omitted. The remaining inequalities (6b) and (6d) are linearized as in (7), by replacing \(\max_k \{\hat{\lambda}_j^k\}\) by \(\rho_j\), \(\min_k \{\hat{\lambda}_j^k\}\) by \(\sigma_j\), and incorporating the additional inequalities \(\rho_j \geq \hat{\lambda}_j^k\) and \(\sigma_j \leq \hat{\lambda}_j^k\), for all \(j, k\).

8. If \(\lambda_j > 0\), components \(\alpha_j^k\) are found from equality (5): \(\alpha_j^k = (\lambda_j + \mu_j^k)/\lambda_j\). If \(\lambda_j = 0\) but \(\mu_j^k \neq 0\)
for at least one $k$, equality (5) cannot be satisfied. As follows from the proof of Lemma 5, any such DMU $(X,Y)$ in (6) is a limit point of a sequence of DMUs of the other types stated in Remark 1.

9. In this case inequalities (6b) and (6e) are removed from conditions (6), and the superscript $k$ is dropped. Conditions (6) become the standard statement of the HRS technology (Podinovski 2004a) by the substitution $\mu_j = \hat{\mu}_j - \hat{\nu}_j$, where $\hat{\mu}_j, \hat{\nu}_j \geq 0$, $\forall j \in J$.

10. For example, the incorporation of an upper bound on the proportion $p^k_j$ in program (9) in the form $p^k_j \leq \bar{p}^k_j$, where $\bar{p}^k_j > 0$ is a constant, is equivalent to the incorporation of the linear inequality $\xi^k_j \leq \bar{p}^k_j (\sum_{i \in S} v_i X_{ji})$ in the linear program (13) shown in Appendix C.


12. The CRS and MHRS technologies extend the VRS technology in different ways. The CRS technology allows full scaling of DMUs which is disallowed by the MHRS technology with nonproportional inputs or outputs. On the other hand, the MHRS technology allows selective scaling of component processes, which is impossible in the CRS technology.

13. We use the term “indeterminate” when, for at least one sample of size $n$ out of the 100 replications, the estimated efficiency of all DMUs is equal to 1. (This implies the model does not provide sufficient discrimination.) In such cases the rank correlation for the given sample is indeterminate and so is the average rank correlation.

14. Model (LP-5) is given in the online supplement to the paper of Cherchye et al. (2013).
References


Brief Author Biographies

Victor V. Podinovski is professor of operational research at the School of Business and Economics and co-director of the Centre for Productivity and Performance at Loughborough University, United Kingdom. His research primarily concerns mathematical models of efficiency and productivity analysis, including nonparametric methodologies of data envelopment analysis (DEA).

Ole Bent Olesen is professor of management science at Department of Business and Economics at the University of Southern Denmark, Denmark. His research interests include mathematical and statistical modeling of production economics with focus on productivity and efficiency analysis, in particular data envelopment analysis.

Cláudia S. Sarrico is an associate professor at ISEG Lisbon School of Economics & Management, Universidade de Lisboa, and senior researcher at CIPES Centre for Research in Higher Education Policies in Portugal. Her research focuses on performance measurement and management in education, higher education and science.