Disaggregation for efficiency analysis*

Barnabé Walheer†

January 27, 2019

Abstract

Efficiency analysis is a popular technique used to compare the performances of Decision Making Units (DMUs). Procedures to aggregate DMU-level efficiency and related efficiency concepts to obtain group-specific counterparts have been proposed recently. In this paper, we suggest procedures for the opposite direction: disaggregate DMU-level efficiency and related efficiency concepts into output-specific counterparts. For being largely applicable, based on economic optimization behaviour, easy to use and to interpret, these procedures establish a novel toolkit for efficiency analysis practitioners.
Keywords: Efficiency analysis; disaggregation; Farrell efficiency; scale efficiency; congestion.

*We thank the Editor Victor Podinovski, the Associate Editor, and the two anonymous referees for their very productive comments that substantially improved the paper.
†HEC Liège, Université de Liège, Belgium. email: barnabe.walheer@uliege.be.

1 Introduction

Efficiency analysis, which aims at comparing the performances of Decision Making Units (DMUs), has become popular both as a research instrument and a practical decision-support tool. To define efficiency, there are two coexisting approaches: one structural and the other technical. The former starts from an economic optimization behaviour (such as cost minimization or profit maximization) and characterizes inefficiency in terms of deviation from this economic model.[1] The latter, rather than assuming any economic optimization behaviour, compares the input-output performance and measures efficiency as the distance to the frontier of the production technology.[2] Note that both approaches are related and can, in some conditions, yield the same efficiency results.[3]

There has been much effort recently to find procedures to aggregate (DMU-level) structural and technical efficiency. The main objective is to obtain comparable indicators at the group-level that are non-trivial, fulfil certain properties, and based on non ad-hoc aggregation procedures (with, if possible, an economic interpretation). See, for example, Färe and Zelenyuk (2003, 2005, 2007), Färe, Grosskopf, and Zelenyuk (2004), Kuosmanen, Cherchye, and Sipilainen (2006), Li and Cheng (2007), Nesterenko and Zelenyuk (2007), Pachkova (2009), ten Raa (2011), Peyrache (2013), Färe and Karagiannis (2014), and Walheer (2018e). Attention has not only been given to aggregate efficiency, but also to aggregate other related efficiency concepts: see, for example, Zelenyuk (2006), and Mayer and Zelenyuk (2014) for the Malmquist productivity index, Färe and Zelenyuk (2012) for scale elasticities, Zelenyuk (2015) for scale efficiency, Färe and Karagiannis (2017) for capacity utilization and input congestion, Walheer (2018a) for meta-frontier technology gap, and Walheer (2018e) for scale efficiency and congestion. To derive their aggregation procedures, they had to make certain assumptions on aggregate technology, optimization behaviour, prices, and/or make use of particular aggregation schemes.

While the literature on aggregation procedures is quite extensive, surprisingly few works have considered the opposite direction: disaggregate (DMU-level) efficiency and related efficiency concepts to obtain output-specific indicators. For multi-output DMUs, this is particularly relevant since more standard efficiency analysis models only provide results at the DMU-level and are therefore not discriminant enough. Indeed, for those DMUs, knowing in which outputs they perform better or worst is valuable information. To our knowledge,

[1] For early works on this approach, refer to Afriat (1972), Hanoch and Rothschild (1972), Diewert and Parkan (1983) and Varian (1984).

[2] See, for example, Färe, Grosskopf and Lovell (1994), Cooper, Seiford and Zhu (2004), Coelli, Rao, O'Donnell, and Battese (2005), Cooper, Seiford and Tone (2007), Fried, Lovell and Schmidt (2008), and Cook and Seiford (2009) for reviews on the two approaches.

[3] See Banker and Maindiratta (1988) and Färe and Grosskopf (1995) for the interrelationship between the two. Farrell (1957) relates them by using the concept of allocative efficiency. See Section 3 for more detail.

only few papers, whether directly or not, have considered the disaggregation of (DMU-level) efficiency and related efficiency concepts. Cherchye et al (2013) have defined a cost minimization model for multi-output DMUs and (indirectly) have obtained a disaggregation of the (DMU-level) cost efficiency by means of output-specific cost shares. Cherchye, De Rock, and Walheer (2016) have extended this research in a profit context, and hence (indirectly) obtained a disaggregation of the (DMU-level) profit efficiency by means of output-specific profit shares. In a macroeconomic context, Walheer (2016a, b) has (indirectly) obtained a disaggregation of the country-level (output-oriented) technical efficiency into sector-level technical efficiency by using the output shares for the weights. Finally, Walheer (2018c, d) and Walheer and Zhang (2018) have shown how to decompose several productivity indexes into output-specific counterparts. To obtain their (direct or indirect) disaggregation results, they have modelled each output separately using output-specific production technologies by allocating the DMU-level input vector to each output production process.[4]

This paper variously contributes to the disaggregation for efficiency analysis. Firstly, rather than modelling the production process by means of separated technologies for each output, we consider groups of outputs. In a sense, indeed, modelling each output individually is too extreme since it ignores the natural links between certain types of outputs, such as discretionary and non-discretionary (Banker and Morey (1986)), good and bad (Färe and Grosskopf (2004)), separable and non-separable (Cooper, Seiford, and Tone (2007)). Next, we do not impose specific ways of allocating the inputs to the group of outputs. Assuming particular ways could limit our exposition and, in the worst case, make the method useless for some applications. As a result, our framework fits with several efficiency analysis models, from the more standard models to those assuming particular internal structures of the production process (i.e. particular way of allocating the inputs). Therefore, our framework should be particularly attractive to empirical researchers who are familiar with more standard approaches, as well as to those using more advanced models. Moreover, we do not only focus our attention on the disaggregation of efficiency, but show how to apply the disaggregation principle by considering scale efficiency and congestion. Finally, we use the disaggregation procedures to obtain useful decompositions of the structural efficiency concepts. Thus, we dare think that this demonstration of the benefits and practical use of the disaggregation for efficiency analysis equips practitioners with a new toolkit.

The rest of the paper is structured as follows. Section 2 demonstrates how to disaggregate efficiency and related efficiency concepts in cost minimizing settings. Section 3 explains

[4]Their allocation procedure also fits with efficiency analysis models integrating information on the internal production structure. See, for example, Salerian and Chan (2005), Despic,, Despic, and Paradi (2007), Cherchye, De Rock, and Vermeulen (2008), Färe and Grosskopf (2000), Färe, Grosskopf and Whittaker (2007), Tone and Tsutsui (2009), Cherchye, De Rock, and Walheer (2015), Ding et al (2017), Silva (2018), and Walheer (2018d); and Remark 1 in Section 4.

how to decompose the cost concepts into technical and allocative counterparts. Section 4, in conclusion, makes some remarks.

2 Disaggregation in cost minimizing settings

The starting point of an efficiency analysis is the observation of a sample of DMUs. We consider that every DMU produces Q groups of outputs.[5] What distinguishes the disaggregation procedure is that it considers output-specific production technologies. (Note that it implies that each output-specific technology can contain more than one output). So, instead of considering the production technology of the Q groups of outputs simultaneously, we define the technology for each group of outputs separately. This representation allows us to disaggregate cost efficiency and related cost efficiency concepts into output-specific counterparts, without requiring extra assumptions on any aspects of the production process.

In particular, let $\mathbf{y}^q$ be the output vector of the q-th group, $\mathbf{x}^q$ be the inputs used to produce that output, and $\mathbf{w}^q$ be the associated input price vector. (Note that we do not require the input vectors to be of equal length). We assume that the inputs and their prices are observed. In fact, it is not a strong assumption for many applications. Nevertheless, if the inputs and/or their prices are partially or not observed, the method can still be applied. Section 4 presents an extensive discussion of these cases (see Remarks 2 and 3).

In what follows, we explain how to disaggregate efficiency and related efficiency concepts in cost minimizing settings. In particular, we start by defining cost efficiency and related cost efficiency concepts at the output-specific level. We then do the same at the DMU-level and show how the DMU-level definitions can be disaggregated into their output-specific counterparts. As a final remark, it is straightforward to generalize our disaggregation procedures in cost settings for revenue or profit maximizing settings.[6]

2.1 Output-specific framework

Our first step is to define the technology for each group of outputs. Next, we define our notion of efficiency in cost minimization settings, and derive related efficiency concepts: scale efficiency and congestion. Attractively, these output-specific definitions make it possible to compare DMUs for each output individually. Note that we base our account on well-known definitions of the related efficiency concepts, but it is important to remark that the following

[5]Note that our modelling is applicable to standard models as well as those assuming a particular internal structure of the production process. See Remark 1 in Section 4 for more detail.

[6]Inspirations could be found in Cherchye, De Rock, and Walheer (2016) that have defined profit efficiency (and directional distance function) in a similar context. See also Walheer and Zhang (2018) for dynamic profit efficiency settings.

could be extended with alternative definitions. Our aim is to show how disaggregation procedures work for efficiency analysis.

Define the technology. We define the technology by means of production possibility sets. In particular for output group q, it is given by

$$T^q = \{(\mathbf{x}^q, \mathbf{y}^q) \mid \mathbf{x}^q \text{ can produce } \mathbf{y}^q\}. \tag{1}$$

In words, the set T^q contains the inputs that can produce outputs for group q. As we are interested in the cost (or input) side of the production process, we could alternatively define the technology by means of input requirement sets, defined as follows:

$$I^q(\mathbf{y}^q) = \{\mathbf{x}^q \mid (\mathbf{x}^q, \mathbf{y}^q) \in T^q\}. \tag{2}$$

We impose some regularity conditions on the input requirement sets captured by Axioms 1-4. In fact, while these Axioms are similar to those stated in Färe and Primont (1995), they are adapted here to our output-specific setting. Besides being very general, they are common to many popular efficiency analysis models and form an empirically attractive minimal set of assumptions. Finally, note that, in general, cost efficiency evaluation does not require Axioms 3 and 4 (see, for example, Varian (1984) and Tulkens (1993) for discussion). We impose those extra conditions to match the technical counterparts discussed in Section 3.

Axiom 1 (observability means feasibility): $(\mathbf{y}^1, \ldots, \mathbf{y}^Q, \mathbf{x}^1 \ldots, \mathbf{x}^Q)$ is observed $\implies$ $\forall q : \mathbf{x}^q \in I^q(\mathbf{y}^q)$.

Axiom 2 (strong disposability of outputs): $\mathbf{y}^q \geq \mathbf{y}^{q'} \implies I^q(\mathbf{y}^q) \subseteq I^q(\mathbf{y}^{q'})$.

Axiom 3 (strong disposability of inputs): $\mathbf{x}^q \in I^q(\mathbf{y}^q)$ and $\mathbf{x}^{q'} \geq \mathbf{x}^q \implies \mathbf{x}^{q'} \in I^q(\mathbf{y}^q)$.

Axiom 4 (convex input sets): $\mathbf{x}^q \in I^q(\mathbf{y}^q)$ and $\mathbf{x}^{q'} \in I^q(\mathbf{y}^q) \implies \forall \lambda \in [0,1] : \lambda \mathbf{x}^q + (1-\lambda)\mathbf{x}^{q'} \in I^q(\mathbf{y}^q)$.

Axiom 1 says that what is observed is certainly feasible. Or, if we observe $(\mathbf{y}^1, \ldots, \mathbf{y}^Q, , \mathbf{x}^1 \ldots, \mathbf{x}^Q)$, then these inputs can definitively produce the observed outputs. This implies that what we observe does not suffer from measurement error or other statistical noise. Axiom 2 says that, if $\mathbf{x}^q$ can produce $\mathbf{y}^q$, then it can also produce less output. Essentially, this axiom implies that the input requirement sets are nested. Next, Axiom 3 says that more inputs never reduce the outputs and implies that the input requirement sets are monotone. Axioms 2 and 3 also imply that marginal rates of substitution/transformation (between inputs, outputs and inputs and outputs) are nowhere negative or, in other words, that there is no congestion. Finally, Axiom 4 states that, if $\mathbf{x}^q$ and $\mathbf{x}^{q'}$ can produce $\mathbf{y}^q$, any convex combination can also produce the same output.[7] It entails non-decreasing marginal rates of input substitution.[8]

Farrell efficiency. The starting point of the cost efficiency evaluation is the minimal cost for each output q:

$$C^q(\mathbf{y}^q, \mathbf{w}^q) = \min_{\mathbf{x}^q \in I^q(\mathbf{y}^q)} \mathbf{w}^{q'}\mathbf{x}^q. \tag{3}$$

$C^q(\mathbf{y}^q, \mathbf{w}^q)$ selects the minimal input vector, in the input requirement set $I^q(\mathbf{y}^q)$, to produce the output quantity $\mathbf{y}^q$ given the input prices $\mathbf{w}^q$. $C^q(\mathbf{y}^q, \mathbf{w}^q) \leq \mathbf{w}^{q'}\mathbf{x}^q$, and $C^q(\mathbf{y}^q, \mathbf{w}^q) = \mathbf{w}^{q'}\mathbf{x}^q$ means that outputs in group q are produced with minimal cost while $C^q(\mathbf{y}^q, \mathbf{w}^q) < \mathbf{w}^{q'}\mathbf{x}^q$ reflects potential cost savings.

Following Farrell (1957), we define cost efficiency as the ratio of the minimal to the actual cost. Adapting his definition to our specific setting, we obtain for each q:

$$CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q) = \frac{C^q(\mathbf{y}^q, \mathbf{w}^q)}{\mathbf{w}^{q'}\mathbf{x}^q}. \tag{4}$$

As $C^q(\mathbf{y}^q, \mathbf{w}^q) \leq \mathbf{w}^{q'}\mathbf{x}^q$, we obtain a cost efficiency measurement smaller than one. When $C^q(\mathbf{y}^q, \mathbf{w}^q) = \mathbf{w}^{q'}\mathbf{x}^q$, $CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q) = 1$ reflecting cost efficient behaviour, and when $C^q(\mathbf{y}^q, \mathbf{w}^q) < \mathbf{w}^{q'}\mathbf{x}^q$, $CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q) < 1$, meaning that cost could be saved on the production of the outputs in group q. The efficient level of input vector $\mathbf{x}^q$ is therefore given by: $CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)\mathbf{x}^q$.

Scale efficiency. Our previous four technology axioms do not imply any particular structure in terms of returns-to-scale on the output-level production technology, captured by the input requirement set. That is, we remain as general as possible, and thus implicitly assume that the output-specific production processes exhibit variable returns-to-scale. To test formally for scale efficiency, we first have to define our cost efficiency concept under the hypothetical assumption of constant returns-to-scale. This is captured by the following axiom:

Axiom 5 (constant returns-to-scale technology): $\forall k \in \mathbb{R}_0^+ : \mathbf{x}^q \in I^q(\mathbf{y}^q) \implies k\mathbf{x}^q \in I^q(k\mathbf{y}^q)$.

[7] Note that here we adopt a so-called relaxed convexity approach in the sense that we only require the input requirement sets to be convex (see Petersen (1990) and Bogetoft (1996)) and not the production possibility sets, but it is straightforward to adapt the following with the stronger assumption of convex production possibility sets. See, for example, Walheer (2018b).

[8] For formal definitions of T^q and $I^q(\mathbf{y}^q)$, see Petersen (1990) and Bogetoft (1996), and Cherchye et al (2013) and Cherchye, De Rock, and Walheer (2016).

Axiom 5 says that if $\mathbf{x}^q$ can produce $\mathbf{y}^q$, $k\mathbf{x}^q$ can produce $k\mathbf{y}^q$, with k restricted to be strictly positive. Let us denote $\widehat{T}^q$ as the production possibility set satisfying Axioms 1-5, and the corresponding input set by $\widehat{I}^q(\mathbf{y}^q)$.[9] It is straightforward to define the minimal cost and the cost efficiency measurement with respect to those sets. In fact, it suffices to use $\widehat{I}^q(\mathbf{y}^q)$ instead of $I^q(\mathbf{y}^q)$ in (3), we obtain:

$$\widehat{C}^q(\mathbf{y}^q, \mathbf{w}^q) = \min_{\mathbf{x}^q \in \widehat{I}^q(\mathbf{y}^q)} \mathbf{w}^{q\prime}\mathbf{x}^q, \tag{5}$$

$$\widehat{CE}^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q) = \frac{\widehat{C}^q(\mathbf{y}^q, \mathbf{w}^q)}{\mathbf{w}^{q\prime}\mathbf{x}^q}. \tag{6}$$

The interpretation of $\widehat{CE}^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)$ is analogous to the interpretation of $CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)$, the only difference is that the (in)efficient behaviour is evaluated when assuming constant returns-to-scale (Axiom 5). Note also, that by definition $\widehat{CE}^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q) \leq CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)$. It reflects that the reconstruction of the technology under constant returns-to-scale is, in general, greater than the reconstruction of the technology under variable returns-to-scale, or in other words, that $I^q(\mathbf{y}^q)$ is included in $\widehat{I}^q(\mathbf{y}^q)$.

Färe and Grosskopf (1985) define cost scale efficiency as the ratio of the cost efficiency under the hypothetical constant returns-to-scale assumption and the cost efficiency under variable returns-to-scale:

$$CSE^q(\mathbf{y}^q, \mathbf{w}^q) = \frac{\widehat{CE}^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)}{CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)} = \frac{\widehat{C}^q(\mathbf{y}^q, \mathbf{w}^q)}{C^q(\mathbf{y}^q, \mathbf{w}^q)}. \tag{7}$$

As $\widehat{CE}^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q) \leq CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)$, we have that $CSE^q(\mathbf{y}^q, \mathbf{w}^q) \leq 1$. A value of one indicates cost scale efficiency behaviour. When $CSE^q(\mathbf{y}^q, \mathbf{w}^q) < 1$, it reveals cost scale inefficiency, which could be due to decreasing or increasing returns-to-scale.[10] As a final remark, note that $CSE^q(\mathbf{y}^q, \mathbf{w}^q)$ does not depend on input level since $\widehat{CE}^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)$ and $CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)$ have, by definition, the same denominators (see (4) and (6)).

Congestion. As explained earlier, Axioms 2 and 3 imply that marginal rates of substitution/transformation (between inputs, outputs and inputs and outputs) are nowhere negative. In other words, Axioms 2 and 3 imply that there is no congestion in the production process. As such, to capture congestion, we define a weaker version of these axioms. In particular, to define output congestion, we rely on the weak disposability of outputs,

[9] $\widehat{T}^q$ is directly related to T^q, since $\widehat{T}^q = \{\lambda(\mathbf{x}^q, \mathbf{y}^q) \in T^q, \forall \lambda > 0\}$. For formal definitions of $\widehat{T}^q$ and $\widehat{I}^q(\mathbf{y}^q)$, see Petersen (1990) and Bogetoft (1996), and Walheer (2018b).

[10] In practice, it is enough to evaluate cost efficiency to non-increasing returns-to-scale and compare this efficiency measurement to $CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)$. If they are equal, cost scale inefficiency is due to decreasing returns-to-scale. Otherwise, cost scale inefficiency is due to increasing returns-to-scale.

captured by the following axiom:

Axiom 6 (weak disposability of outputs): $\forall \delta \in [0,1] : I^q(\mathbf{y}^q) \subseteq I^q(\delta \mathbf{y}^q)$.

Axiom 6 implies that if $\mathbf{x}^q$ can produce $\mathbf{y}^q$, then $\mathbf{x}^q$ can produce $\delta \mathbf{y}^q$. As such, Axiom 6 is similar to Axiom 2, but with the extra constraint that the outputs must be rescaled proportionally by the same factor δ. Therefore, Axiom 6 is a weaker version of Axiom 2. Let $\widetilde{T}^q$ be the production possibility set that satisfies axioms 1, 3, 4, and 6. The corresponding input set is given by $\widetilde{I}^q(\mathbf{y}^q)$.[11] The corresponding output-specific minimal cost and cost efficiency measurement based on those sets are given by:

$$\widetilde{C}^q(\mathbf{y}^q, \mathbf{w}^q) = \min_{\mathbf{x}^q \in \widetilde{I}^q(\mathbf{y}^q)} \mathbf{w}^{q\prime} \mathbf{x}^q. \tag{8}$$

$$\widetilde{CE}^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q) = \frac{\widetilde{C}^q(\mathbf{y}^q, \mathbf{w}^q)}{\mathbf{w}^{q\prime} \mathbf{x}^q}. \tag{9}$$

Again the interpretation of $\widetilde{CE}^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)$ is analogous to $CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)$, but with respect to the assumption of weak disposability of outputs. As noticed before, stronger assumptions/axioms on the technology can only decrease efficiency. As such, we have that $\widetilde{CE}^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q) \geq CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)$ since weak disposability is less restrictive than strong disposability.

Färe and Grosskopf (1983) defines output congestion as the ratio of (technical) efficiency with respectively strong and weak output disposability. Adapting their definition to our specific setting, we obtain:

$$COC^q(\mathbf{y}^q, \mathbf{w}^q) = \frac{CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)}{\widetilde{CE}^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)} = \frac{C^q(\mathbf{y}^q, \mathbf{w}^q)}{\widetilde{C}^q(\mathbf{y}^q, \mathbf{w}^q)}. \tag{10}$$

The cost output congestion measurement is thus bounded from above by one since, as explained previously, $\widetilde{CE}^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q) \geq CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)$. Unity means that there is no congestion of the outputs, while smaller values imply more congestions. As $CSE^q(\mathbf{y}^q, \mathbf{w}^q)$, $COC^q(\mathbf{y}^q, \mathbf{w}^q)$ does not depend on input level since $CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)$ and $\widetilde{CE}^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)$ have the same denominators.

2.2 DMU-level framework

In this Section, we show how to disaggregate the (DMU-level) efficiency and related efficiency concepts in cost minimization settings. In other words, we show how to obtain the

[11] Note that we adopt here a variable-returns-to-scale setting, there are no reasons not to do the same under a constant returns-to-scale assumption (i.e. by adding Axiom 5). Formal definitions of these sets can easily be obtained by modifying the setting considered in Podinovski and Kuosmanen (2011) to our specific setting.

(DMU-level) efficiency and related efficiency concepts from the output-specific counterparts. Specially, let $CE(\mathbf{y}, \mathbf{x}, \mathbf{w})$, $CSE(\mathbf{y}, \mathbf{w})$, and $COC(\mathbf{y}, \mathbf{w})$ be, respectively, the (DMU-level) Farrell cost efficiency, scale efficiency, and output congestion indicators; where we denote, to simply our notations: $\mathbf{y} = (\mathbf{y}^1, \ldots, \mathbf{y}^Q)$, $\mathbf{x} = (\mathbf{x}^1, \ldots, \mathbf{x}^Q)$, and $\mathbf{w} = (\mathbf{w}^1, \ldots, \mathbf{w}^Q)$.[12] These DMU-level definitions are important for multi-output DMUs since, contrary to the output-specific definitions that are focused on the output-specific production level ($\mathbf{y}^q$), they provide efficiency information for the DMU-level production ($\mathbf{y}$).

Our aim, therefore, is to find such disaggregation schemes as to we keep a natural way to interpret the DMU-level indicators, i.e. they fulfil desirable properties, and such that the disaggregation procedures remain natural and compatible with economic interpretation. In particular, let F_1, F_2, and F_3 be the functions that relates the (DMU-level) efficiency concepts to their output-specific counterparts:

$$CE(\mathbf{y}, \mathbf{x}, \mathbf{w}) = F_1\left(CE^1(\mathbf{y}^1, \mathbf{x}^1, \mathbf{w}^1), \ldots, CE^Q(\mathbf{y}^Q, \mathbf{x}^Q, \mathbf{w}^Q)\right). \tag{11}$$

$$CSE(\mathbf{y}, \mathbf{w}) = F_2\left(CSE^1(\mathbf{y}^1, \mathbf{w}^1), \ldots, CSE^Q(\mathbf{y}^Q, \mathbf{w}^Q)\right). \tag{12}$$

$$COC(\mathbf{y}, \mathbf{w}) = F_3\left(COC^1(\mathbf{y}^1, \mathbf{w}^1), \ldots, COC^Q(\mathbf{y}^Q, \mathbf{w}^Q)\right). \tag{13}$$

In fact, as will be shown later, all disaggregation schemes will be obtained by linear weightings. This is particularly attractive since it facilitates the practical use and interpretation of disaggregation procedures. To facilitate our following demonstration, let us define, for every $\zeta^q \in \mathbb{R}$, the following weights:

$$\alpha^q(\mathbf{x}, \mathbf{w}; \zeta^q) = \frac{\mathbf{w}^{q'}(\mathbf{x}^q \zeta^q)}{\sum_{q=1}^{Q} \mathbf{w}^{q'}(\mathbf{x}^q \zeta^q)}. \tag{14}$$

The weights present some interesting properties. Firstly, the weights $\alpha^q(\mathbf{x}, \mathbf{w}; \zeta^q)$ depend only on the input $\mathbf{x}$ and their price $\mathbf{w}$ (and ζ^q). That is, they depend on all the output-specific inputs and their associated prices , but not on the output quantities. This is not surprising since the goal is to aggregate output-specific cost efficiency and related cost efficiency concepts into DMU-level counterparts. Next, the weights are non-negative and sum the unity: $\sum_{q=1}^{Q} \alpha^q(\mathbf{x}, \mathbf{w}; \zeta^q) = \sum_{q=1}^{Q} \frac{\mathbf{w}^{q'}(\mathbf{x}^q \zeta^q)}{\sum_{q=1}^{Q} \mathbf{w}^{q'}(\mathbf{x}^q \zeta^q)} = \frac{\sum_{q=1}^{Q} \mathbf{w}^{q'}(\mathbf{x}^q \zeta^q)}{\sum_{q=1}^{Q} \mathbf{w}^{q'}(\mathbf{x}^q \zeta^q)} = 1$. Finally, they have a nice economic interpretation: $\alpha^q(\mathbf{x}, \mathbf{w}; \zeta^q)$ represents the cost share of output q when the inputs are corrected/rescaled by the factor ζ^q. Note that, if the inputs are not corrected ($\zeta^q = 1, \forall q$), the weights correspond to the output-specific cost shares; and if the inputs

[12] Note that, $\mathbf{y}$, $\mathbf{x}$, and $\mathbf{w}$ are not the standard definitions of the outputs, the inputs, and the input prices at the DMU-level but are here matrices. The difference is due to our consideration of an output-specific setting. Nevertheless, it is easy to relate our setting to more standard efficiency analysis models. In fact, these models are a particular case of ours. See Remark 1 in Section 4 for a discussion.

are corrected by cost efficiency ($\zeta^q = CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q), \forall q$), the weights correspond to the output-specific minimal cost shares. More extensive discussions on those weights will also be provided below (See also Remark 7 in Section 4).

Farrell efficiency. To obtain our Farrell cost efficiency at the DMU-level, we first have to define the actual cost and minimal costs at that level. Attractively, both costs can easily be linked to output-specific actual and minimal costs. In particular, it suffices to sum the output-specific actual and minimal costs to obtain the DMU-level equivalences. This is quite intuitive and in line with Farrell (1957) and more recent works aggregation and disaggregation procedures. Indeed, a similar result is obtained for the aggregation procedure by, for example, Färe and Zelenyuk (2003, 2005, 2007), Färe, Grosskopf and Zelenyuk (2004), Zelenyuk (2006, 2015), Färe and Karagiannis (2017), and Walheer (2018a, e); and for the disaggregation procedure by, for example, Cherchye et al (2013), Cherchye, De Rock, and Walheer (2016), and Walheer (2018b, c). We provide a formal proof in the Appendix; also, refer to Remark 1 in Section 4 for examples and related discussion. More precisely, we obtain, $\sum_{q=1}^{Q} \mathbf{w}^{q\prime}\mathbf{x}^q$ and $\sum_{q=1}^{Q} C^q(\mathbf{y}^q, \mathbf{w}^q)$ for the DMU-level actual and minimal costs, respectively.

Clearly, the property of the DMU-level costs are the same as their respective output-specific counterparts. A first observation is that $C^q(\mathbf{y}^q, \mathbf{w}^q) \leq \mathbf{w}^{q\prime}\mathbf{x}^q$ for all q, imply that $\sum_{q=1}^{Q} C^q(\mathbf{y}^q, \mathbf{w}^q) \leq \sum_{q=1}^{Q} \mathbf{w}^{q\prime}\mathbf{x}^q$, i.e. the minimal cost is bounded by the actual cost. Next, if each output is produced with minimal costs, i.e. $C^q(\mathbf{y}^q, \mathbf{w}^q) = \mathbf{w}^{q\prime}\mathbf{x}^q$ for all q, then $\sum_{q=1}^{Q} C^q(\mathbf{y}^q, \mathbf{w}^q) = \sum_{q=1}^{Q} \mathbf{w}^{q\prime}\mathbf{x}^q$, i.e. the actual cost coincides with the minimal cost revealing cost efficiency. Finally, if at least one output group is produced inefficiently, i.e. $C^q(\mathbf{y}^q, \mathbf{w}^q) < \mathbf{w}^{q\prime}\mathbf{x}^q$ for at least one q, we have that $\sum_{q=1}^{Q} C^q(\mathbf{y}^q, \mathbf{w}^q) < \sum_{q=1}^{Q} \mathbf{w}^{q\prime}\mathbf{x}^q$, revealing cost inefficiency. Therefore, it is enough to produce one output group with a non-minimal cost to reveal cost inefficient behaviour. This shows the high discriminatory power of the output-specific modelling but also reveals that (DMU-level) efficiency measurements are, in a sense, too strong and do not give enough detailed results on the inefficient behaviour (i.e. the source(s) of inefficiency).

As we did before for the output-specific framework, we can define Farrell cost efficiency as the ratio of the minimal cost and the actual cost:

$$CE(\mathbf{y}, \mathbf{x}, \mathbf{w}) = \frac{C(\mathbf{y}, \mathbf{w})}{\sum_{q=1}^{Q} \mathbf{w}^{q\prime}\mathbf{x}^q} = \frac{\sum_{q=1}^{Q} C^q(\mathbf{y}^q, \mathbf{w}^q)}{\sum_{q=1}^{Q} \mathbf{w}^{q\prime}\mathbf{x}^q}. \tag{15}$$

As such, $CE(\mathbf{y}, \mathbf{x}, \mathbf{w})$ depends only on output-level cost minimizing conditions. As noticed before $\sum_{q=1}^{Q} C^q(\mathbf{y}^q, \mathbf{w}^q) \leq \sum_{q=1}^{Q} \mathbf{w}^{q\prime}\mathbf{x}^q$, making $CE(\mathbf{y}, \mathbf{x}, \mathbf{w})$ bounded from above by one, with unity reveals cost efficiency while smaller value means greater cost inefficiency. As a result, the property of $CE(\mathbf{y},\mathbf{x},\mathbf{w})$ are exactly the same as those of the output-specific cost efficiency $CE^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q)$.

While, in (15), $CE(\mathbf{y},\mathbf{x},\mathbf{w})$ is exclusively linked to output-specific cost concepts, it does not match the disaggregation procedure defined in (11), i.e. a function F_1 that relates $CE(\mathbf{y},\mathbf{x},\mathbf{w})$ to the output-specific cost efficiencies $CE(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q)$ only. Still, it is possible to use (15) to find this disaggregation procedure. In particular, it suffices to multiply and divide (15) by $\mathbf{w}^{q'}\mathbf{x}^q$. Rearranging the terms gives:

$$CE(\mathbf{y},\mathbf{x},\mathbf{w}) = \sum_{q=1}^{Q} \frac{\mathbf{w}^{q'}\mathbf{x}^q}{\sum_{q=1}^{Q}\mathbf{w}^{q'}\mathbf{x}^q}\frac{C^q(\mathbf{y}^q,\mathbf{w}^q)}{\mathbf{w}^{q'}\mathbf{x}^q}. \tag{16}$$

As the first term is $\alpha^q(\mathbf{x},\mathbf{w};1)$ (i.e. $\zeta^q = 1, \forall q$: no correction or rescale has been done to the inputs), and the second term is the output-specific cost efficiency defined in (4), we obtain our first disaggregation procedure (F_1):

$$CE(\mathbf{y},\mathbf{x},\mathbf{w}) = \sum_{q=1}^{Q} \alpha^q(\mathbf{x},\mathbf{w};1)\, CE^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q). \tag{17}$$

The weights $\alpha^q(\mathbf{x},\mathbf{w};1)$ have a nice economic interpretation. They could be interpreted as the share of the total budget allocated to output group q, and therefore allow us to identify which output-specific cost efficiency contributes more to the cost efficiency.[13]

Scale efficiency. Let use denote $\widehat{CE}(\mathbf{y},\mathbf{x},\mathbf{w})$ as the (DMU-level efficiency) under the hypothetical assumption that the output-specific technologies exhibit constant returns-to-scale. Using the relationship established before for the (DMU-level) cost efficiency measurement (see (17)), we directly obtain the following relationship:

$$CSE(\mathbf{y},\mathbf{w}) = \frac{\widehat{CE}(\mathbf{y},\mathbf{x},\mathbf{w})}{CE(\mathbf{y},\mathbf{x},\mathbf{w})} = \frac{\sum_{q=1}^{Q}\alpha^q(\mathbf{x},\mathbf{w};1)\widehat{CE}^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q)}{\sum_{q=1}^{Q}\alpha^q(\mathbf{x},\mathbf{w};1)\, CE^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q)}. \tag{18}$$

Note that the weights are similar for the cost efficiency measurements under the variable and constant returns-to-scale assumptions. In fact, this is intuitive since the weights are independent from the reconstructed production technology: they represent the output-specific cost shares. Also, from (7), we have $\widehat{CE}^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q) \leq CE^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q)$. Multiplying by the output-specific cost share $\alpha^q(\mathbf{x},\mathbf{w};1)$ (non-negative by definition) and summing over

[13] The weights $\alpha^q(\mathbf{x},\mathbf{w};1)$ generalize the weights in the disaggregation procedures of Cherchye et al (2013), Cherchye, De Rock, and Walheer (2016), and Walheer (2018b, c); and are consistent with the weights found for aggregate procedures by, for example, Färe and Zelenyuk (2003, 2005, 2007), Färe, Grosskopf and Zelenyuk (2004), Zelenyuk (2006, 2015), Färe and Karagiannis (2017), and Walheer (2018a, e).

the Q outputs will not affect the inequality. We obtain: $\sum_{q=1}^{Q} \alpha^q(\mathbf{x}, \mathbf{w}; 1)\widehat{CE}^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)$ $\leq \sum_{q=1}^{Q} \alpha^q(\mathbf{x}, \mathbf{w}; 1)\, CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)$, making $CSE(\mathbf{y}, \mathbf{w}) \leq 1$. Clearly, $CSE(\mathbf{y}, \mathbf{w}) = 1$ reflects scale efficiency of the Q groups of outputs, while a value smaller than one reveals scale inefficiency (the source(s) of scale inefficiency could be found by looking at the output-specific scale efficiency measurements $CSE^q(\mathbf{y}^q, \mathbf{w}^q)$). Finally, as $CSE^q(\mathbf{y}^q, \mathbf{w}^q)$, $CSE(\mathbf{y}, \mathbf{w})$ does not depend on the output levels. All in all, the properties of the scale efficiency measurement $CSE(\mathbf{y}, \mathbf{w})$ are the same as those at the output-level.

Building on (18), we can obtain our second disaggregation procedure (i.e. F_2). It suffices to multiply top and bottom by $\alpha^q(\mathbf{x}, \mathbf{w}; 1)\, CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)$. Rearranging the terms gives:

$$CSE(\mathbf{y}, \mathbf{w}) = \sum_{q=1}^{Q} \frac{\alpha^q(\mathbf{x}, \mathbf{w}; 1)\, CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)}{\sum_{q=1}^{Q} \alpha^q(\mathbf{x}, \mathbf{w}; 1)\, CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)} \frac{\widehat{CE}^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)}{CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)}. \tag{19}$$

In fact the weights $\frac{\alpha^q(\mathbf{x}, \mathbf{w}; 1)\, CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)}{\sum_{q=1}^{Q} \alpha^q(\mathbf{x}, \mathbf{w}; 1)\, CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)}$ coincide with $\alpha^q(\mathbf{x}, \mathbf{w}; CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q))$, i.e. the output-specific cost shares rescaled by $CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)$:

$$\begin{aligned}
\frac{\alpha^q(\mathbf{x}, \mathbf{w}; 1)\, CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)}{\sum_{q=1}^{Q} \alpha^q(\mathbf{x}, \mathbf{w}; 1)\, CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)} &= \frac{\frac{\mathbf{w}^{q\prime}\mathbf{x}^q}{\sum_{q=1}^{Q} \mathbf{w}^{q\prime}\mathbf{x}^q}\, CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)}{\sum_{q=1}^{Q} \frac{\mathbf{w}^{q\prime}\mathbf{x}^q}{\sum_{q=1}^{Q} \mathbf{w}^{q\prime}\mathbf{x}^q}\, CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)}, &(20)\\
&= \frac{\mathbf{w}^{q\prime}(\mathbf{x}^q\, CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q))}{\sum_{q=1}^{Q} \mathbf{w}^{q\prime}(\mathbf{x}^q\, CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q))}, &(21)\\
&= \alpha^q(\mathbf{x}, \mathbf{w}; CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)). &(22)
\end{aligned}$$

Interestingly, we can equally rewrite these weights as follows (by replacing $CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)$ by its definition in (21)):

$$\alpha^q(\mathbf{x}, \mathbf{w}; CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)) = \frac{C^q(\mathbf{y}^q, \mathbf{w}^q)}{\sum_{q=1}^{Q} C^q(\mathbf{y}^q, \mathbf{w}^q)}. \tag{23}$$

Thus,these weights can also be interpreted as the output-specific minimal cost share.

Using the weights $\alpha^q(\mathbf{x}, \mathbf{w}; CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q))$ and by noting that the second term in (21) corresponds to the output-specific cost scale efficiency (see (7)), we obtain our second disaggregation procedure (F_2):

$$CSE(\mathbf{y}, \mathbf{w}) = \sum_{q=1}^{Q} \alpha^q(\mathbf{x}, \mathbf{w}; CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q))\, CSE^q(\mathbf{y}^q, \mathbf{w}^q). \tag{24}$$

Congestion. Just like to scale efficiency, we define output congestion at the DMU-level as follows:

$$COC(\mathbf{y},\mathbf{w}) = \frac{CE(\mathbf{y},\mathbf{x},\mathbf{w})}{\widetilde{CE}(\mathbf{y},\mathbf{x},\mathbf{w})} = \frac{\sum_{q=1}^{Q} \alpha^q(\mathbf{x},\mathbf{w};1)\, CE^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q)}{\sum_{q=1}^{Q} \alpha^q(\mathbf{x},\mathbf{w};1)\, \widetilde{CE}^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q)}. \tag{25}$$

where $\widetilde{CE}(\mathbf{y},\mathbf{x},\mathbf{w})$ represents the DMU-level cost efficiency under the assumption that the output-specific input requirement sets are defined under weak disposability.

In fact, it suffices to follow the same step as for the scale efficiency in order to obtain the disaggregation procedure (i.e. F_3). We obtain:

$$COC(\mathbf{y},\mathbf{w}) = \sum_{q=1}^{Q} \frac{\alpha^q(\mathbf{x},\mathbf{w};1)\widetilde{CE}^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q)}{\sum_{q=1}^{Q} \alpha^q(\mathbf{x},\mathbf{w};1)\widetilde{CE}^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q)} \frac{CE^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q)}{\widetilde{CE}^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q)}, \tag{26}$$

$$= \sum_{q=1}^{Q} \alpha^q(\mathbf{x},\mathbf{w};\widetilde{CE}^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q))\, COC^q(\mathbf{y}^q,\mathbf{w}^q). \tag{27}$$

Alternatively, the weights again can be redefined as follows:

$$\alpha^q(\mathbf{x},\mathbf{w};\widetilde{CE}^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q)) = \frac{\widetilde{C}^q(\mathbf{y}^q,\mathbf{w}^q)}{\sum_{q=1}^{Q} \widetilde{C}^q(\mathbf{y}^q,\mathbf{w}^q)}. \tag{28}$$

It turns that the weights $\alpha^q(\mathbf{x},\mathbf{w};\widetilde{CE}^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q))$ have a nice economic interpretation: they represent the output-specific minimal cost share with respect to the technologies when assuming weak disposability.

Clearly, $COC(\mathbf{y},\mathbf{w})$ fulfils the desirable properties. This directly comes from the relationships between $CE^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q)$ and $\widetilde{CE}^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q)$ explained in (11). As noticed previously for $CSE(\mathbf{y},\mathbf{w})$, multiplying by the output-specific cost share (non-negative by definition) and summing over Q outputs will not affect the inequality. As such, $COC(\mathbf{y},\mathbf{w})$ is bounded from above by one, with one indicating no congestion for the Q groups of outputs. A value smaller than one indicates congestion and the source(s) can be identified using the $COC^q(\mathbf{y}^q,\mathbf{w}^q)$. Finally, as $COC^q(\mathbf{y}^q,\mathbf{w}^q)$, $COC(\mathbf{y},\mathbf{w})$ does not depend on the input level; and the weights are similar for $CE^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q)$ and $\widetilde{CE}^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q)$ in (25). This is explained, as noticed for (18), by the independence of the weights to the technology axioms imposed on the production technology when considering cost efficiency.

3 Decomposition

In this Section, we show that the disaggregation procedure for the structural efficiency approach can be used to define a technical and allocative efficiency measurement for the DMU. This defines a useful decomposition of the DMU-level structural efficiency. Interestingly, we can also obtain such decomposition for the cost-based scale efficiency and congestion concepts. As such, this Section complements the results for the structural efficiency approach and increases the usefulness of the disaggregation procedures for practitioners.

Contrary to the structural efficiency approach, the technical approach does not start from an economic optimization behaviour. Instead, it compares the input-output performances and therefore only requires the observation of $\mathbf{y}^q$ and $\mathbf{x}^q$ for every q. As such, there is no reason that structural efficiency coincides with technical efficiency. Nevertheless, Farrell (1957) shows the two types of efficiency are related by the concept of allocative efficiency (inefficiency due to non-optimal allocation of inputs/outputs given the prices). In fact, structural efficiency can always be obtained as a product of technical and allocative efficiency. This also applies to the related efficiency concepts. Building on this relationship, we use our disaggregation schemes to obtain technical and allocative counterparts. As in Section 2, we focus our demonstration on the input-oriented technical and allocative counterpart; the though is easily extended for other orientations.[14]

In particular, let us use $TE(\mathbf{y}, \mathbf{x}, \mathbf{w})$, $TSE(\mathbf{y}, \mathbf{x}, \mathbf{w})$, and $TOC(\mathbf{y}, \mathbf{x}, \mathbf{w})$ to denote the DMU-level Farrell efficiency, scale efficiency, and congestion indicators in the technical context; and $AE(\mathbf{y}, \mathbf{x}, \mathbf{w})$, $ASE(\mathbf{y}, \mathbf{x}, \mathbf{w})$, and $AOC(\mathbf{y}, \mathbf{x}, \mathbf{w})$ their allocative counterparts. More formally, we use functions to represent the connections with the output-specific definitions:

$$TE(\mathbf{y}, \mathbf{x}, \mathbf{w}) = F_4\left(TE^1(\mathbf{y}^1, \mathbf{x}^1), \ldots, TE^Q(\mathbf{y}^Q, \mathbf{x}^Q)\right). \tag{29}$$

$$TSE(\mathbf{y}, \mathbf{x}, \mathbf{w}) = F_5\left(TSE^1(\mathbf{y}^1, \mathbf{x}^1), \ldots, TSE^Q(\mathbf{y}^Q, \mathbf{x}^Q)\right). \tag{30}$$

$$TOC(\mathbf{y}, \mathbf{x}, \mathbf{w}) = F_6\left(TOC^1(\mathbf{y}^1, \mathbf{x}^1), \ldots, TOC(\mathbf{y}^Q, \mathbf{x}^Q)\right). \tag{31}$$

$$AE(\mathbf{y}, \mathbf{x}, \mathbf{w}) = F_7\left(AE^1(\mathbf{y}^1, \mathbf{x}^1, \mathbf{w}^1), \ldots, AE^Q(\mathbf{y}^Q, \mathbf{x}^Q, \mathbf{w}^Q)\right), \tag{32}$$

$$ASE(\mathbf{y}, \mathbf{x}, \mathbf{w}) = F_8\left(ASE^1(\mathbf{y}^1, \mathbf{x}^1, \mathbf{w}^1), \ldots, ASE^Q(\mathbf{y}^Q, \mathbf{x}^Q, \mathbf{w}^Q)\right). \tag{33}$$

$$AOC(\mathbf{y}, \mathbf{x}, \mathbf{w}) = F_9\left(AOC^1(\mathbf{y}^1, \mathbf{x}^1, \mathbf{w}^1), \ldots, AOC(\mathbf{y}^Q, \mathbf{x}^Q, \mathbf{w}^Q)\right). \tag{34}$$

where an index with a postscript q represents the output-specific counterparts (for formal definitions see Section 2). Note that the DMU-level technical indicators depend on the input prices while their output-specific counterparts do not. This becomes clearer in the following.

[14] Inspirations could be found in Cherchye, De Rock, and Walheer (2016) and Walheer and Zhang (2018) that have defined directional distance function in a similar context.

Farrell efficiency. The definition of (input-oriented) technical efficiency goes back to to Farrell (1957). Inspired by Debreu (1951), he defines (input-oriented) technical efficiency as the maximal equiproportionate/radial input reduction that still allows for the production of a certain output quantity.[15] Adapting Farrell's definition to our specific setting, we obtain the following:

$$TE^q(\mathbf{y}^q, \mathbf{x}^q) = \inf\{\theta \mid \theta \mathbf{x}^q \in I^q(\mathbf{y}^q)\}. \tag{36}$$

Generally, $TE^q(\mathbf{y}^q, \mathbf{x}^q)$ is smaller than 1, and a lower value indicates greater technical inefficiency. Also, note that $TE^q(\mathbf{y}^q, \mathbf{x}^q)$, contrary to $CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)$, does not depend on the input price vector $\mathbf{w}^q$. This directly follows from their respective definitions. Farrell (1957) noticed that there is no particular reason why cost efficiency should coincide with technical efficiency, but found a natural way to relate the two types of efficiency:

$$CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q) = TE^q(\mathbf{y}^q, \mathbf{x}^q) AE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q). \tag{37}$$

$AE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)$ stands for (input-oriented) allocative efficiency and measures inefficiency due to non-optimal allocation of inputs given the input prices $\mathbf{w}^q$.[16] As such, the two efficiency measurements coincide when $AE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q) = 1$, or in other words, when there is no inappropriate allocation of inputs. This occurs, in particular, when the input prices are the most favourable ones (i.e. the shadow prices), giving an interpretation of technical efficiency as a (shadow) cost efficiency.[17] Therefore, we have: $TE^q(\mathbf{y}^q, \mathbf{x}^q) \geq CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)$ making $AE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)$ smaller than 1.

To obtain F_4 and F_7, we make use of the following decomposition of cost efficiency into technical and allocative counterparts (obtained by combining (17) and (37)):

$$\begin{aligned} CE(\mathbf{y}, \mathbf{x}, \mathbf{w}) &= \sum_{q=1}^{Q} \alpha^q(\mathbf{x}, \mathbf{w}; 1)\, CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q), && (38) \\ &= \sum_{q=1}^{Q} \alpha^q(\mathbf{x}, \mathbf{w}; 1)\, TE^q(\mathbf{y}^q, \mathbf{x}^q) AE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q). && (39) \end{aligned}$$

Multiplying top and bottom by $\sum_{q=1}^{Q} \mathbf{w}^{q'}(\mathbf{x}^q\, TE^q(\mathbf{y}^q, \mathbf{x}^q))$ and rearranging the terms

[15] The (input-oriented) technical efficiency is also the reciprocal of the (input) distance function introduced by Shephard (1953,1970):

$$\frac{1}{TE^q(\mathbf{y}^q, \mathbf{x}^q)} = D(\mathbf{y}^q, \mathbf{x}^q) = \sup\left\{\eta \mid \left(\frac{\mathbf{x}^q}{\eta}\right) \in I^q(\mathbf{y}^q)\right\}. \tag{35}$$

It means that we can alternatively consider distance functions instead of technical efficiency measurements in Section 3.

[16] See Farrell (1957) and Färe and Primont (1995) for more detail.

[17] Formally: $TE^q(\mathbf{y}^q, \mathbf{x}^q) = \max_{\mathbf{w}^q} CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)$. See also Remarks 3 and 5 in Section 4.

gives us:

$$CE(\mathbf{y},\mathbf{x},\mathbf{w}) = \sum_{q=1}^{Q} \alpha^q(\mathbf{x},\mathbf{w};1)\, TE^q(\mathbf{y}^q,\mathbf{x}^q) \sum_{q=1}^{Q} \alpha^q(\mathbf{x},\mathbf{w};\, TE^q(\mathbf{y}^q,\mathbf{x}^q)) AE^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q). \quad (40)$$

Note that the weights for technical efficiency are similar as the weights of cost efficiency (see (17)). As a consequence, the weights depend on the input prices, which can be surprising for a technical concept (see our discussion below). On the contrary, the weights for allocative efficiency are different. In fact, $\alpha^q(\mathbf{x},\mathbf{w};\, TE^q(\mathbf{y}^q,\mathbf{x}^q))$ are simply a modified version of the weights based on cost efficiency with technical efficiency as the rescale factor.

Let us define:

$$TE(\mathbf{y},\mathbf{x},\mathbf{w}) = \sum_{q=1}^{Q} \alpha^q(\mathbf{x},\mathbf{w};1)\, TE^q(\mathbf{y}^q,\mathbf{x}^q). \quad (41)$$

$$AE(\mathbf{y},\mathbf{x},\mathbf{w}) = \sum_{q=1}^{Q} \alpha^q(\mathbf{x},\mathbf{w};\, TE^q(\mathbf{y}^q,\mathbf{x}^q)) AE^q(\mathbf{y}^q,\mathbf{x}^q). \quad (42)$$

As such, (41) and (42) correspond to the functions F_4 and F_7:

$$CE(\mathbf{y},\mathbf{x},\mathbf{w}) = TE(\mathbf{y},\mathbf{x},\mathbf{w}) AE(\mathbf{y},\mathbf{x},\mathbf{w}). \quad (43)$$

$TE(\mathbf{y},\mathbf{x},\mathbf{w})$ and $AE(\mathbf{y},\mathbf{x},\mathbf{w})$ are, respectively, the (input-oriented) technical and allocative efficiency measurements at the DMU-level. As, for every q, $TE^q(\mathbf{y}^q,\mathbf{x}^q)$ is smaller than one, and since the weights $\alpha^q(\mathbf{x},\mathbf{w};1)$ sum to one by definition (see (16)), we end with $TE(\mathbf{y},\mathbf{x},\mathbf{w})$ being also smaller than one. Clearly, one indicates technical efficiency at the DMU-level, while lower values indicate inefficiency. The same reasoning also applies to allocative efficiency: $AE^q(\mathbf{y}^q,\mathbf{x}^q)$ are smaller than one, and the weights $\alpha^q(\mathbf{x},\mathbf{w};\, TE^q(\mathbf{y}^q,\mathbf{x}^q))$ sum to one by definition (see (14)). As such, $AE(\mathbf{y},\mathbf{x},\mathbf{w})$ has the same property as $AE^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q)$.

As a final remark, note that $TE(\mathbf{y},\mathbf{x},\mathbf{w})$, contrary to $TE^q(\mathbf{y}^q,\mathbf{x}^q)$, depends on the input prices (because of the weights). In a sense, $TE(\mathbf{y},\mathbf{x},\mathbf{w})$ is therefore a bad measure of technical efficiency since it does not depend on the inputs and outputs only (See, in an aggregation context, Färe and Zelenyuk (2003), and Remarks 3 and 7 in Section 4). Putting it differently, these measurements are defined in terms of the output-specific counterparts exclusively, and thus lack connection with the DMU's production process (contrary to the cost concepts in Section 2).

Scale efficiency. Just as for cost efficiency, natural definitions of scale efficiency in the technical and allocative contexts are given by the ratio of the efficiency under the hypothetical constant returns-to-scale assumption and the efficiency under variable returns-to-scale:

$$TSE^q(\mathbf{y}^q, \mathbf{x}^q) = \frac{\widehat{TE}^q(\mathbf{y}^q, \mathbf{x}^q)}{TE^q(\mathbf{y}^q, \mathbf{x}^q)}. \tag{44}$$

$$ASE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q) = \frac{\widehat{AE}^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)}{AE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)}. \tag{45}$$

where $\widehat{TE}^q(\mathbf{y}^q, \mathbf{x}^q) = \inf\{\beta \mid \beta\mathbf{x}^q \in \widehat{I}^q(y^q)\}$ is the technical efficiency with respect to $\widehat{I}^q(y^q)$, and $\widehat{AE}^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q) = \frac{\widehat{CE}^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)}{\widehat{TE}^q(\mathbf{y}^q, \mathbf{x}^q)}$. The interpretation of these measurements is analogous to those in the variable returns-to-scale case. (Note that, as for cost efficiency, the weights are the same for constant and variable returns-to-scale). Clearly, $TSE^q(\mathbf{y}^q, \mathbf{w}^q) \leq 1$ since $\widehat{TE}^q(\mathbf{y}^q, \mathbf{x}^q) \leq TE^q(\mathbf{y}^q, \mathbf{x}^q)$, but this is not guaranteed for $ASE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)$. This is not a big issue since the main interest of this part is the technical efficiency concepts; allocative efficiency concepts are only present because they are required to relate cost efficiency concepts to technical efficiency ones. Again, a value of one indicates scale efficiency behaviour. At this point, it is important to notice that the cost-based and technical scale efficiency measurements are related through a special case of homotheticity, the scale-homotheticity. When the technology satisfies the scale homotheticity property, both scale efficiency measurements coincide. We refer to Zelenyuk (2014) for more discussion.

In the spirit of Farrell (1957), we can decompose cost scale efficiency into two parts (see also Färe and Grosskopf (1985)):

$$CSE^q(\mathbf{y}^q, \mathbf{w}^q) = TSE^q(\mathbf{y}^q, \mathbf{x}^q) ASE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q). \tag{46}$$

Building on our previous results in (24), we have:

$$\begin{aligned} CSE(\mathbf{y}, \mathbf{w}) &= \sum_{q=1}^{Q} \alpha^q(\mathbf{x}, \mathbf{w}; CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q))\, CSE^q(\mathbf{y}^q, \mathbf{w}^q), && (47)\\ &= \sum_{q=1}^{Q} \alpha^q(\mathbf{x}, \mathbf{w}; CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q))\, TSE^q(\mathbf{y}^q, \mathbf{x}^q) ASE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q). && (48) \end{aligned}$$

By multiplying top and bottom by $\sum_{q=1}^{Q} \mathbf{w}^{q'}(\mathbf{x}^q\, CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)\, TSE^q(\mathbf{y}^q, \mathbf{x}^q))$ and rearranging the terms, we obtain the following:

$$\begin{aligned} CSE(\mathbf{y},\mathbf{w}) = &\sum_{q=1}^{Q} \alpha^q(\mathbf{x},\mathbf{w};\, CE^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q))\, TSE^q(\mathbf{y}^q,\mathbf{x}^q) \\ &\times \sum_{q=1}^{Q} \alpha^q(\mathbf{x},\mathbf{w};\, CE^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q)\, TSE^q(\mathbf{y}^q,\mathbf{x}^q))\, ASE^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q). \end{aligned} \tag{49}$$

As such, the weights for technical scale efficiency are the same as those for cost scale efficiency. They can be interpreted as the output-specific potential cost shares. The weights for allocative efficiency are different and imply that the inputs are rescaled by both $CE^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q)$ and $TSE^q(\mathbf{y}^q,\mathbf{x}^q)$. Interestingly, we could rewrite those weights as follows:

$$\alpha^q(\mathbf{x},\mathbf{w};\, CE^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q)\, TSE^q(\mathbf{y}^q,\mathbf{x}^q))) = \frac{\mathbf{w}^{q\prime}(\mathbf{x}^q\, CE^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q)\, TSE^q(\mathbf{y}^q,\mathbf{x}^q))}{\sum_{q=1}^{Q} \mathbf{w}^{q\prime}(\mathbf{x}^q\, CE^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q)\, TSE^q(\mathbf{y}^q,\mathbf{x}^q))}, \tag{50}$$

$$= \frac{C^q(\mathbf{y}^q,\mathbf{w}^q)\, TSE^q(\mathbf{y}^q,\mathbf{x}^q)}{\sum_{q=1}^{Q} C^q(\mathbf{y}^q,\mathbf{w}^q)\, TSE^q(\mathbf{y}^q,\mathbf{x}^q)}. \tag{51}$$

That is, $\alpha^q(\mathbf{x},\mathbf{w};\, CE^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q)\, TSE^q(\mathbf{y}^q,\mathbf{x}^q))$ can alternatively be interpreted as the output-specific minimal cost shares rescaled by technical scale efficiency.

We end the decomposition of scale efficiency (i.e. F_5 and F_8) by defining:

$$TSE(\mathbf{y},\mathbf{x},\mathbf{w}) = \sum_{q=1}^{Q} \alpha^q(\mathbf{x},\mathbf{w};\, CE^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q))\, TSE^q(\mathbf{y}^q,\mathbf{x}^q), \tag{52}$$

$$ASE(\mathbf{y},\mathbf{x},\mathbf{w}) = \sum_{q=1}^{Q} \alpha^q(\mathbf{x},\mathbf{w};\, CE^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q)\, TSE^q(\mathbf{y}^q,\mathbf{x}^q))\, ASE^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q). \tag{53}$$

Combining these last two equations with (49), we have:

$$CSE(\mathbf{y},\mathbf{w}) = TSE(\mathbf{y},\mathbf{x},\mathbf{w})\, ASE(\mathbf{y},\mathbf{x},\mathbf{w}). \tag{54}$$

Clearly, $TSE(\mathbf{y},\mathbf{x},\mathbf{w})$ shares the same properties as the output-specific counterparts $TSE^q(\mathbf{y}^q,\mathbf{x}^q)$: they are smaller than one, with one indicating scale efficiency for all the outputs. This comes from the definition of $TSE^q(\mathbf{y}^q,\mathbf{x}^q)$ (they are smaller than one), and from the definition of the weights (they sum to unity). Again, we highlight that $TSE(\mathbf{y},\mathbf{x},\mathbf{w})$ is a bad measurement of technical scale efficiency, but it enters rather naturally in the decomposition of the cost-based counterparts. For $ASE(\mathbf{y},\mathbf{x},\mathbf{w})$, as for $ASE(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q)$, the domain is unknown.

Congestion. The starting point of the technical definition of output congestion is the technical efficiency measurement with respect to the set $\widetilde{I}^q(\mathbf{y}^q)$:

$$\widetilde{TE}^q(\mathbf{y}^q, \mathbf{x}^q) = \inf\{\nu \mid \nu\mathbf{x}^q \in \widetilde{I}^q(\mathbf{y}^q)\}. \tag{55}$$

Following Färe and Grosskopf (1983), we define technical output congestion as

$$TOC^q(\mathbf{y}^q, \mathbf{x}^q) = \frac{TE^q(\mathbf{y}^q, \mathbf{x}^q)}{\widetilde{TE}^q(\mathbf{y}^q, \mathbf{x}^q)}. \tag{56}$$

As for the cost efficiency, more assumptions/axioms can only decrease technical efficiency, yielding to $TE^q(\mathbf{y}^q, \mathbf{x}^q) \leq \widetilde{TE}^q(\mathbf{y}^q, \mathbf{x}^q)$. As such, technical output congestion $TOC^q(\mathbf{y}^q, \mathbf{x}^q)$ is smaller than one, with one indicating no congestion. By defining $\widetilde{AE}^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q) = \frac{\widetilde{CE}^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)}{\widetilde{TE}^q(\mathbf{y}^q, \mathbf{x}^q)}$, we obtain the allocative output congestion indicator:

$$AOC^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q) = \frac{AE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)}{\widetilde{AE}^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)}. \tag{57}$$

As for the allocative scale efficiency, we cannot specify the domain of this indicator.

Again, drawing on Farrell (1957), we can decompose cost output congestion into the two parts defined earlier:

$$COC^q(\mathbf{y}^q, \mathbf{w}^q) = TOC^q(\mathbf{y}^q, \mathbf{x}^q) AOC^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q). \tag{58}$$

Combining the previous equation with equation (27), multiplying top and bottom by $\sum_{q=1}^{Q} \mathbf{w}^{q\prime}(\mathbf{x}^q \widetilde{CE}^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)\, TOC^q(\mathbf{y}^q, \mathbf{x}^q))$, and rearranging the terms, gives us:

$$\begin{aligned} COC(\mathbf{y}, \mathbf{w}) &= \sum_{q=1}^{Q} \alpha^q(\mathbf{x}, \mathbf{w}; \widetilde{CE}^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q))\, TOC^q(\mathbf{y}^q, \mathbf{x}^q) AOC^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q), && (59) \\ &= \sum_{q=1}^{Q} \alpha^q(\mathbf{x}, \mathbf{w}; \widetilde{CE}^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q))\, TOC^q(\mathbf{y}^q, \mathbf{x}^q) \\ &\times \sum_{q=1}^{Q} \alpha^q(\mathbf{x}, \mathbf{w}; \widetilde{CE}^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)\, TOC^q(\mathbf{y}^q, \mathbf{x}^q)) AOC^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q). && (60) \end{aligned}$$

We obtain the connections (i.e. F_6 and F_9) by defining:

$$TOC(\mathbf{y},\mathbf{x},\mathbf{w}) = \sum_{q=1}^{Q} \alpha^q(\mathbf{x},\mathbf{w};\widetilde{CE}^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q))\,TOC^q(\mathbf{y}^q,\mathbf{x}^q). \tag{61}$$

$$AOC(\mathbf{y},\mathbf{x},\mathbf{w}) = \sum_{q=1}^{Q} \alpha^q(\mathbf{x},\mathbf{w};\widetilde{CE}^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q)\,TOC^q(\mathbf{y}^q,\mathbf{x}^q))AOC^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q). \tag{62}$$

Thus we have:

$$COC(\mathbf{y},\mathbf{w}) = TOC(\mathbf{y},\mathbf{x},\mathbf{w})AOC(\mathbf{y},\mathbf{x},\mathbf{w}). \tag{63}$$

From (61), it is clear that $TOC(\mathbf{y},\mathbf{x},\mathbf{w})$ has the same property as $TOC^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q)$, i.e. the indicator is smaller than one, with one indicating no congestion for all the outputs. At this point, we highlight that $TOC(\mathbf{y},\mathbf{x},\mathbf{w})$ depend on the input prices and is therefore a bad measurement of technical congestion. The weights, which coincide with those of the cost output congestion indicator, can be interpreted as the output-specific minimal cost shares with respect to the weak disposability technologies (see (28)). As for $AOC^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q)$, we cannot specify the domain of $AOC(\mathbf{y},\mathbf{x},\mathbf{w})$. The weights for this indicator imply that the inputs are rescaled by both $\widetilde{CE}^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q)$ and $TOC^q(\mathbf{y}^q,\mathbf{x}^q)$. As we rewrote the weights for allocative scale efficiency in (50) – (51), we can rewrite $\alpha^q(\mathbf{y},\mathbf{w};\widetilde{CE}^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q)\,TOC^q(\mathbf{y}^q,\mathbf{x}^q))$ as the output-specific minimal cost shares rescaled by technical output congestion:

$$\alpha^q(\mathbf{x},\mathbf{w};\widetilde{CE}^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q)\,TOC^q(\mathbf{y}^q,\mathbf{x}^q))) = \frac{\mathbf{w}^{q'}(\mathbf{x}^q\widetilde{CE}^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q)\,TOC^q(\mathbf{y}^q,\mathbf{x}^q))}{\sum_{q=1}^{Q}\mathbf{w}^{q'}(\mathbf{x}^q\widetilde{CE}^q(\mathbf{y}^q,\mathbf{x}^q,\mathbf{w}^q)\,TOC^q(\mathbf{y}^q,\mathbf{x}^q))}, \tag{64}$$

$$= \frac{\widetilde{C}^q(\mathbf{y}^q,\mathbf{w}^q)\,TOC^q(\mathbf{y}^q,\mathbf{x}^q)}{\sum_{q=1}^{Q}\widetilde{C}^q(\mathbf{y}^q,\mathbf{w}^q)\,TOC^q(\mathbf{y}^q,\mathbf{x}^q)}. \tag{65}$$

4 Concluding remarks

This Section develops several remarks, each considering a specific topic.

Remark 1: Allocation of inputs. Our output-specific setting naturally allows us to allocate inputs to specific output production process. The techniques considering this option generally model each output separately. We find this type of modelling rather restrictive since it does not give the practitioners the option to take the links between certain types of outputs into account (bad/good, discretionary/non-discretionary, separable/non-separable, etc.). Therefore, techniques modelling each output individually could be seen as a particular case of ours when it is assumed that each group is composed by one output.

Different ways have been suggested to allocate inputs to outputs. Firstly, inputs could be completely allocated to the output-specific technologies. These inputs have been considered in, for example, Färe and Grosskopf (2000), Färe, Grosskopf and Whittaker (2007), Tone and Tsutsui (2009), Cherchye et al (2013), and Walheer (2016a, b). In that case, adding the input quantities ($\mathbf{x}^q$) give the DMU-level inputs. Next, inputs could be used to jointly produce outputs. This input type has been considered in, for example, Cherchye et al (2013), Cherchye, De Rock, and Walheer (2016), Walheer (2017, 2018b, c), and Walheer and Zhang (2018). In that case, the output-specific inputs ($\mathbf{x}^q$) coincide with the DMU-level inputs. Afterwards, inputs could be used to produce a group of outputs, as, for example in Salerian and Chan (2005), Despic et al (2007), and Cherchye, De Rock, and Walheer (2015). In that case, the output-specific inputs $\mathbf{x}^q$ would be different for each group of outputs. Finally, inputs could be proportional to the outputs. These types of input have been considered in, for example, Podinovski (2004a, c, 2009), Podinovski et al (2014), Podinovski and Husai (2017), and Podinovski, Olesen and Sarrico (2018). Clearly, other ways could be chosen, and all these ways to allocate the inputs could also be combined.Rather than imposing any specific ways, our modelling leaves that choice to the practitioners.

As a final remark, note that the chosen way(s) to allocate the inputs to the outputs could also imply some (implicit or explicit) relationships between the output-specific input prices $\mathbf{w}^q$ and the DMU-level input prices (see, for example, Cherchye et al (2008, 2013, 2016) and Walheer (2017, 2018b, c) for examples and discussion).

Remark 2: Observation of the output-specific inputs. Our methodology, like that of most of the previous works cited in Remark 1, is based on the observation of the output-specific inputs $(\mathbf{x}^1, \dots, \mathbf{x}^Q)$. In fact, for many applications, this is not a strong assumption, since DMUs often use systems that explicitly allocate inputs/costs to outputs, as for example Activity Based Costing. Nevertheless, if this information is not (or partially) available, the disaggregation procedure can still be used. Indeed, if only the DMU-level inputs are observed, the allocation over output can be reconstructed. It will only complicate the model and its computation (by, for example, making the model non-linear). See, for example, Cook, Habadou, and Teunter (2000), Beasley (2003), Li, Yang, Liang, and Hua (2009), Yu, Chern, and Hsiao (2013), Du, Cook, Liang, and Zhu (2014), and Walheer (2016b). Refer also to Cherchye et al (2013) who have provided a discussion on how to integrate these techniques in a similar setting to ours.

Remark 3: Observation of the output-specific input prices. Our methodology requires the observation of the output-specific inputs $(\mathbf{w}^1, \dots, \mathbf{w}^Q)$. As discussed in Remark

2, in many applications observing the output-specific inputs is not a strong assumption. As such, the same should hold for their prices. Next, if only the DMU-level input prices are observed, the output-specific input prices can be reconstructed by allocating the inputs to the outputs. Indeed, specific allocation implies relationships between those prices (see, for example, Cherchye et al (2008, 2013, 2016) and Walheer (2017, 2018b, c) for examples). Finally, if no input prices are observed, it is still possible to use the disaggregation procedures by relying on the shadow input prices. These prices are the most favourable ones and therefore give the benefit of the doubt to the DMUs. It is a common procedure in efficiency analysis in the absence of true price information. In particular, the shadow cost efficiency would be defined as:

$$SCE^q(\mathbf{y}^q, \mathbf{x}^q) = \max_{\mathbf{w}^q} CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q). \tag{66}$$

When the output-specific input prices are computed, they could be used in the weights $\alpha^q(\mathbf{x}^q, \mathbf{w}^q; \zeta^q)$. As a final remark, note that the realism of the computed output-specific input prices can be increased by choosing appropriate intervals and/or by linking those prices to the DMU-level input prices (see, for example, Cherchye et al (2008, 2013, 2016) and Walheer (2017, 2018b, c) for examples).

Remark 4: Practical implementation. Our measurement can be computed using both parametric (e.g. Stochastic Frontier Analysis) or nonparametric methods (e.g. Data Envelopment Analysis). One of the advantages of Data Envelopment Analysis methods is that efficiency is very easily computed by solving simple linear programs. In fact, it suffices to compute the output-specific cost or technical efficiency measurements to obtain all the weights and DMU-level counterparts. For the sake of compactness and given our very general setting, we do not give the linear programs here but refer to the papers of Cherchye et al (2013) for $CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)$, and to Walheer (2018) for $\widehat{CE}^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)$. $\widetilde{CE}^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)$ is easily obtained by imposing weak disposability in the programs of $CE^q(\mathbf{y}^q, \mathbf{x}^q, \mathbf{w}^q)$ (see also Podinovski and Kuosmanen (2011)). For technical efficiency counterparts: $TE^q(\mathbf{y}^q, \mathbf{x}^q)$, $\widehat{TE}^q(\mathbf{y}^q, \mathbf{x}^q)$, and $\widetilde{TE}^q(\mathbf{y}^q, \mathbf{x}^q)$, see also Cherchye, De Rock, and Walheer (2015). Note that these papers consider more restrictive settings but it is straightforward to generalize their linear programs to alternative settings. Note also that the multi-output setting offers the option to investigate for outliers in every output direction (see Walheer (2019)).

Remark 5: DMU-level technical efficiency. In a similar context, another technical efficiency measurement has been suggested by Cherchye et al (2013) and Cherchye, De Rock, and Walheer (2015), defined in our extended setting, as follows:

$$TE^u(\mathbf{y}, \mathbf{x}) = \inf\{\gamma \mid \forall q : \gamma \mathbf{x}^q \in I^q(\mathbf{y}^q)\}, \tag{67}$$

In fact, their alternative technical measurement could be seen as an upper bound of ours: $TE^u(\mathbf{y},\mathbf{x}) \geq TE(\mathbf{y},\mathbf{x},\mathbf{w})$. To see this, it is enough to rewrite $TE^u(\mathbf{y},\mathbf{x})$ as follows (this comes from the common factor γ in (67)):

$$TE^u(\mathbf{y},\mathbf{x}) = \max\{TE^1(\mathbf{y}^1,\mathbf{x}^1),\ldots,TE^Q(\mathbf{y}^Q,\mathbf{x}^Q)\}. \tag{68}$$

As such, $TE^u(\mathbf{y},\mathbf{x})$ only picks the maximal value of the output-specific technical efficiency measurements. In other words, $TE^u(\mathbf{y},\mathbf{x})$ puts a weight of one to the maximal output-specific technical efficiency measurement.[18] Instead, our measurement $TE(\mathbf{y},\mathbf{x},\mathbf{w})$ takes all output-specific measurement into account while weighting by their output-specific cost share. As a final remark, note that an advantage of $TE^u(\mathbf{y},\mathbf{x})$ is that it does not depend on the input prices, but the weighting is too extreme.

Remark 6: Combining aggregation and disaggregation. In this paper, we have shown how to obtain the DMU-level efficiency and related efficiency concepts from the output-specific counterparts. These DMU-level definitions could be used in the aggregation procedures of, for example, Färe and Zelenyuk (2003, 2005, 2007, 2012), Färe, Grosskopf, and Zelenyuk (2004), Kuosmanen, Cherchye, and Sipilainen (2006), Zelenyuk (2006, 2015), Li and Cheng (2007), Nesterenko and Zelenyuk (2007), Pachkova (2009), ten Raa (2011), Peyrache (2013), Färe and Karagiannis (2014, 2017), Mayer and Zelenyuk (2014), and Walheer (2018a, e), to obtain group-specific indicators. As such, we would obtain a group-specific indicator while starting from output-specific optimization conditions, using a double weighting system (from output to DMU-level, and from DMU-level to group-level).

Remark 7: One input case. In the case where one input is used to produce the outputs, the weight would become:

$$\alpha^q(x^q,w^q;\zeta^q) = \frac{w^q(x^q\zeta^q)}{\sum_{q=1}^{Q} w^q(x^q\zeta^q)}. \tag{69}$$

If, in addition, we assume that the output-specific input prices are equal: $w^q = w$, which is not a strong assumption in one input case, we obtain:

$$\alpha^q(x^q,w;\zeta^q) = \frac{w(x^q\zeta^q)}{\sum_{q=1}^{Q} w(x^q\zeta^q)} = \frac{(x^q\zeta^q)}{\sum_{q=1}^{Q}(x^q\zeta^q)}. \tag{70}$$

That is, the weights are price-independent. This is also a solution to the non-observation of those prices (see Remark 3).

[18] In the same spirit, we could define a lower bound for our technical measurement $TE(\mathbf{y},\mathbf{x},\mathbf{w})$ as: $TE^l(\mathbf{y},\mathbf{x}) = \min\{TE^1(\mathbf{y}^1,\mathbf{x}^1),\ldots,TE^Q(\mathbf{y}^Q,\mathbf{x}^Q)\}$.

Note that if we also assume that outputs are produced by the same input quantity and that no rescaling is needed, the weights become the arithmetic average:

$$\alpha^q(x, w; 1) = \frac{(x1)}{\sum_{q=1}^{Q}(x1)} = \frac{1}{Q}. \tag{71}$$

This shows that arithmetic average is probably a bad choice for weighting in efficiency analysis since, in fact, it requires strong assumptions on the inputs and input prices.

Remark 8: Extensions. Naturally, different types of extension could be proposed. Firstly, the disaggregation procedure could be extended to more efficiency and related efficiency concepts, and to alternative definitions of the presented concepts. A second natural extension would be to consider the possibility of reallocation amongst output production processes. In other words, verifying whether the DMU could benefit from reallocating the input quantities across outputs. Reallocation of inputs between DMUs has been considered by Nesterenko and Zelenyuk (2007) and Mayer and Zelenyuk (2014) for the static and dynamic settings, respectively.

Next, the disaggregation procedure could be used for indexes. Indexes consider the dynamic changes of efficiency, while we have focused our attention to static measurements. A very popular index is the Malmquist productivity index, introduced by Malmquist (1953), and used in the production context by Caves, Christensen, and Diewert (1982). It is well-known that the Malmquist productivity index can be decomposed into different indexes to better understand the changes in efficiency (see Färe et al (1994)). Therefore, suggesting a disaggregation procedure for this index and its decomposition could lead to valuable information for multi-output DMUs. The same clearly holds true for indicators. Note that a first attempt has been proposed by Walheer (2018d) for the Malmquist index, Walheer (2018b) for the cost-based Malmquist index, and Walheer and Zhang (2018) for the Luenberger indicator and the Malmquist-Luenberger index. Finally, other weights could be suggested for the disaggregation procedure. Here, we suggest a linear weighting procedure. Without claiming that the disaggregation procedure is unique or should be linear, the advantages of our procedure are its easy use and interpretation (based on economic optimization behaviour). The major disadvantage is probably that the technical concept depends on the input prices, and are defined in terms of the output-specific counterparts exclusively.

References

[1] Afriat S., 1972, "Efficiency Estimation of Production Functions", *International Economic Review* 13, 568-598.

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

[3] Banker R. D., Maindiratta A., 1988, "Nonparametric Analysis of Technical and Allocative Efficiencies in Production", *Econometrica* 56, 1315-1332.

[4] Banker R. D., Morey, R., 1986, "Efficiency analysis for exogenously fixed inputs and outputs", *Operations Research* 34 (4), 513-521.

[5] Beasley J. E., 2003, "Allocating fixed costs and resources via data envelopment analysis", *European Journal of Operational Research* 147, 198-216.

[6] Bogetoft P., 1996, "DEA on relaxed convexity assumptions", *Management Science* 42, 457-465.

[7] Caves D. W., Christensen L. R., Diewert W. E., 1982, "The economic theory of index numbers and the measurement of input, output and productivity", *Econometrica* 50, 1393-1414.

[8] Cherchye L., De Rock B., Dierynck B., Roodhooft F., Sabbe J., 2013, "Opening the Black Box of Efficiency Measurement: Input Allocation in Multi-Output Settings", *Operations Research* 61, 1148-1165.

[9] Cherchye L., De Rock B., Vermeulen F., 2008, "Cost-Efficient Production Behavior Under Economies of Scope: A Nonparametric Methodology", *Operations Research* 56, 204-221.

[10] Cherchye L., De Rock B., Walheer B., 2015, "Multi-output efficiency with good and bad outputs", *European Journal of Operational Research* 240, 872-881.

[11] Cherchye L., De Rock B., Walheer B., 2016, "Multi-Output Profit Efficiency And Directional Distance Functions", *Omega* 61, 100-109.

[12] Coelli T. J., Rao D. S. P., O'Donnell C. J., Battese G. E., 2005, "An introduction to efficiency and productivity analysis", *Springer*.

[13] Cook W. D., Hababou M., Tuenter H. J. H., 2000, "Multicomponent efficiency measurement and shared inputs in data envelopment analysis: An application to sales and service performance in bank branches", *Journal of Productivity Analysis* 14, 209-224.

[14] Cook W. D., Seiford L. M., 2009, "Data Envelopment Analysis (DEA) - Thirty years on", *European Journal of Operational Research* 192, 1-17.

[15] Cooper W. W., Seiford L. M., Zhu J., 2004, "Handbook on Data Envelopment Analysis, Second Edition", *Springer Edition.*

[16] Cooper W. W., Seiford L. M., Tone K., 2007, "Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software, Second Edition", *Springer.*

[17] Debreu G., 1951, "The Coefficient of Resource Utilization", *Econometrica* 19(3), 273-292.

[18] Despic O., Despic M., Paradi J., 2007, "DEA-R: ratio-based comparative efficiency model, its mathematical relation to DEA and its use in applications", *Journal of Productivity Analysis* 28, 33-44.

[19] Diewert W. E., Parkan C., 1983, "Linear Programming Tests of Regularity Conditions for Production Frontiers", *Quantitative Studies on Production and Prices.*

[20] Ding J., Dong W., Liang L., Zhu J., 2017, "Goal congruence analysis in multi-Division Organizations with shared resources based on data envelopment analysis", *European Journal of Operational Research* 263, 961-973.

[21] Du J., Cook W. D., Liang L., Zhu J., 2014, "Fixed cost and resource allocation based on DEA cross-efficiency", *European Journal of Operational Research* 235, 206-214.

[22] Färe R., Grosskopf S., 1983, "Measuring Congestion in Production", *Journal of Economics* 43, 257-271.

[23] Färe R., Grosskopf S., 1985, "A Nonparametric Cost Approach to Scale Efficiency", *The Scandinavian Journal of Economics* 87, 594-604.

[24] Färe R., Grosskopf S., 2000, "Network DEA", *Socio-Economic Planning Sciences* 34, 35-49.

[25] Färe R., Grosskopf S., Lovell C. A. K., 1994, "Production Frontier", *Cambridge University Press.*

[26] Färe R., Grosskopf S., Norris M., Zhang Z., 1994, "Productivity growth, technical progress and efficiency change in industrialized countries", *American Economic Review*, 84, 66-83.

[27] Färe R., Grosskopf S., 1995, "Non-Parametric Tests of Regularity, Farrell Efficiency, and Goodness of Fit", *Journal of Econometrics* 69, 415-425.

[28] Färe R., Grosskopf S., Norris M., 1997, "Productivity growth, technical progress and efficiency change in industrialized countries: Reply", *American Economic Review*, 87, 1040-1043.

[29] Färe R., Grosskopf S., 2004, "Modelling undesirable factors in efficiency evaluation: Comment", *European Journal of Operational Research*, 157, 242-245.

[30] Färe R., Grosskopf S., Whittaker G., 2007, "Network DEA", *Modeling Data Irregularities and Structural Complexities in Data Envelopment Analysis, J. Zhu and W. Cook, Eds. Springer.*

[31] Färe R., Grosskopf S., Zelenyuk V., 2004, "Aggregation of Cost Efficiency: Indicators and Indexes Across Firms", *Academia Economic Papers* 32(3), 395-411.

[32] Färe R., Karagiannis G., 2014, "A postscript on aggregate Farrell efficiencies", *European Journal of Operational Research* 233, 784-786.

[33] Färe R., Karagiannis G., 2017, "The denominator rule for share-weighting aggregation", *European Journal of Operational Research* 260, 1175-1180.

[34] Färe R., Primont D., 1995, "Multi-output Production and Duality: Theory and Applications", *Kluwer Academic Publishers* Boston.

[35] Färe R., Zelenyuk V., 2003, "On aggregate Farrell efficiencies", *European Journal of Operational Research* 146, 615-620.

[36] Färe R., Zelenyuk V., 2005, "On Farrell's Decomposition and Aggregation", *International Journal of Business and Economics* 4(2), 167-171.

[37] Färe R., Zelenyuk V., 2007, "Extending Färe and Zelenyuk (2003)", *European Journal of Operational Research* 179, 594-595.

[38] Färe R., Zelenyuk V., 2012, "Aggregation of scale elasticities across firms", *Applied Economics Letters* 19, 1593-1597.

[39] Farrell M., 1957, "The Measurement of Productive Efficiency", *Journal of the Royal Statistical Society* Series 1, General, 120, Part 3, 253-281.

[40] Fried H., Lovell C. A .K., Schmidt S., 2008, "The Measurement of Productive Efficiency and Productivity Change", *Oxford University Press.*

[41] Hanoch G., Rothschild M., 1972, "Testing Assumptions of Production Theory: A Nonparametric Approach", *Journal of Political Economy* 80, 256-275.

[42] Kuosmanen T. , Cherchye L. , Sipilainen T., 2006, "The law of one price in data envelopment analysis: restricting weight flexibility across firms", *European Journal of Operational Research* 170, 735-757 .

[43] Li S. K. , Cheng Y. S., 2007, "Solving the puzzles of structural efficiency", *European Journal of Operational Research* 180 , 713-722.

[44] Li Y. J., Yang F., Liang L., Hua Z. S., 2009, "Allocating the fixed cost as a complement of other cost inputs: A DEA approach.", *European Journal of Operational Research* 197, 389-401.

[45] Malmquist S., 1953, "Index numbers and indifference surfaces", *Trabajos de Estatistica* 4, 209-242.

[46] Maniadakis N., Thanassoulis E., 2004, "A cost Malmquist productivity index", *European Journal of Operational Research* 154, 396-409.

[47] Mas-Colell M., Whinston D., Green J. R., 1995, "Microeconomic Theory", *Oxford University Press* Oxford.

[48] Mayer A., Zelenyuk V., 2014, "Aggregation of Malmquist productivity indexes allowing for reallocation of resources", *European Journal of Operational Research* 238, 774-785.

[49] Nesterenko, V. , Zelenyuk, V. 2007, "Measuring potential gains form reallocation of resources", *Journal of Productivity Analysis* 28 , 107-116 .

[50] Pachkova E. V., 2009, "Restricted reallocation of resources", *European Journal of Operational Research* 196 , 1049-1057.

[51] Petersen N. C., 1990, "Data envelopment analysis on a relaxed set of assumptions", *Management Science* 36, 305-314.

[52] Podinovski V. V., 2009, "Production technologies based on combined proportionality assumptions ", *Journal of Productivity Analysis* 32, 21-26.

[53] Podinovski V. V., 2018, "Returns to scale in convex production technologies", *European Journal of Operational Research* 258, 970-982.

[54] Podinovski V. V., Husain W. R. W., 2017, "The hybrid returns-to-scale model and its extension by production trade-offs: an application to the efficiency assessment of public universities in Malaysia ", *Annals Of Operations Research* 250, 65-84.

[55] Podinovski V. V., Ismail I., Bouzdine-Chameeva T., Zhang W., 2014, "Combining the assumptions of variable and constant returns to scale in the efficiency evaluation of secondary schools", *European Journal of Operational Research* 239, 504-513.

[56] Podinovski V. V., Kuosmanen T., 2011, "Modelling weak disposability in data envelopment analysis under relaxed convexity assumptions", *European Journal of Operational Research* 211, 577-585.

[57] Podinovski V. V., Olesen O. B., Sarrico S. C., 2018, "Nonparametric production technologies with multiple component processes", *Operations Research* 66, 282-300.

[58] Salerian, J. and Chan C., 2005, "Restricting multiple-output multiple-input DEA models by disaggregating the output-input vector", *Journal of Productivity Analysis* 24, 5-29.

[59] Shephard R. W., 1953, "Cost and Production Functions", *Princeton University Press.*

[60] Shephard R. W., 1970, "Theory of Cost and Production Functions", *Princeton University Press.*

[61] Silva M. C. A., 2018, "Output-specific inputs in DEA: An application to courts of justice in Portugal", *Omega* 79, 43-53.

[62] ten Raa T., 2011, "Benchmarking and industry performance", *Journal of Productivity Analysis* 36 , 285-292.

[63] Tone K., Tsutsui M., 2009, "Network DEA: A slacks-based measure approach", *European Journal of Operational Research* 197, 243-252.

[64] Tulkens H., 1993, "On FDH analysis: Some methodological issues and applications to retail banking, courts and urban transit", *Journal of Productivity Analysis* 4, 183-210.

[65] Varian H. R., 1984, "The Non-Parametric Approach to Production Analysis", *Econometrica* 52, 579-598.

[66] Walheer B., 2016a, "A multi-sector nonparametric production-frontier analysis of the economic growth and the convergence of the European countries", *Pacific Economic Review* 21(4), 498-524.

[67] Walheer B., 2016b, "Growth and Convergence of the OECD countries: A Multi-Sector Production-Frontier Approach", *European Journal of Operational Research* 252, 665-675.

[68] Walheer B., 2018a, "Aggregation of metafrontier technology gap ratios: the case of European sectors in 1995-2015", *European Journal of Operational Research* 269, 1013-1026.

[69] Walheer B., 2018b, "Cost Malmquist productivity index: an output-specific approach for group comparison", *Journal of Productivity Analysis* 49(1), 79-94.

[70] Walheer B., 2018c, "Disaggregation of the Cost Malmquist Productivity Index with joint and output-specific inputs", *Omega* 75, 1-12.

[71] Walheer B., 2018d, "Malmquist productivity index for multi-output producers: An application to electricity generation plants", *Socio-Economic Planning Sciences* doi.org/10.1016/j.seps.2018.02.003

[72] Walheer B., 2018e, Scale, congestion, and technical efficiency of European countries: a sector-based nonparametric approach", *Empirical Economics* doi.org/10.1007/s00181-018-1426-7

[73] Walheer B., 2019, "Input allocation in multi-output settings: nonparametric robust efficiency measurements", *Journal of the Operational Research Society* doi.org/10.1080/01605682.2018.1516175

[74] Walheer B., Zhang L.-J. 2018, "Profit Luenberger and Malmquist-Luenberger indexes for multi-activity decision making units: the case of the star-rated hotel industry in China", *Tourism Management* 69, 1-11.

[75] Yu M.-M., Chern C.-C., Hsiao B., 2013, "Human resource rightsizing using centralized data envelopment analysis: Evidence from Taiwan's airports", *Omega* 41(1), 119-130.

[76] Zelenyuk V., 2006, "Aggregation of Malmquist productivity indexes", *European Journal of Operational Research* 174, 1076-1086.

[77] Zelenyuk V., 2014, "Scale efficiency and homotheticity: equivalence of primal and dual measures", *Journal of Productivity Analysis* 42, 15-24.

[78] Zelenyuk V., 2015, "Aggregation of scale efficiency", *European Journal of Operational Research* 240, 269-277.

Appendix

The proof is inspired by the proof for the aggregation case provided by Färe, Grosskopf, and Zelenyuk (2004), which is itself a particular version of the proof of the aggregation of profit function of Mas-Colell, Whinston and Green (1995).

A first step is to define the technology at the DMU-level. We impose very weak conditions for that technology set. The only requirement is that the input-output combinations are feasible for all output-specific technologies. Formally, we obtain:

$$T = \{(\mathbf{x}, \mathbf{y}) \mid \forall q \in \{1, \ldots, Q\} : (\mathbf{x}^q, \mathbf{y}^q) \in T^q\}. \tag{72}$$

In words, $\mathbf{x}$ can produce $\mathbf{y}$ if, and only if, every output quantity $\mathbf{y}^q$ can be produced by $\mathbf{x}^q$. In a similar vein, we define the DMU-level input requirement set as follows:

$$I(\mathbf{y}) = \{\mathbf{x} \mid \forall q \in \{1, \ldots, Q\} : \mathbf{x}^q \in I^q(\mathbf{y}^q)\}. \tag{73}$$

Next, we proof that the minimal cost at the DMU-level, denoted by $C(\mathbf{y}, \mathbf{w})$, is obtained as the sum of the output-specific minimal costs: $C(\mathbf{y}, \mathbf{w}) = \sum_{q=1}^{Q} C^q(\mathbf{y}^q, \mathbf{w}^q)$. Note that we assume that the DMU-level actual cost is defined by the sum of the output-specific actual costs, i.e. $\sum_{q=1}^{Q} \mathbf{w}^{q\prime}\mathbf{x}^q$. This holds true for several input types (see Remark 1). The proof contains two steps:

1. For $q \in \{1, \ldots, Q\}$, let $\mathbf{x}^q \in I^q(\mathbf{y}^q)$ be arbitrary, then $\mathbf{x} \in I(\mathbf{y})$ by (73). By construction, minimal cost at the DMU-level cannot exceed actual cost: $C(\mathbf{y}, \mathbf{w}) \leq \sum_{q=1}^{Q} \mathbf{w}^{q\prime}\mathbf{x}^q$. As $\mathbf{x}^q$ for $q \in \{1, \ldots, Q\}$ are arbitrary, we can choose the cost minimizer levels, which gives:

$$C(\mathbf{y}, \mathbf{w}) \leq \sum_{q=1}^{Q} C^q(\mathbf{y}^q, \mathbf{w}^q). \tag{74}$$

2. Let $\mathbf{x} \in I(\mathbf{y})$ be arbitrary, then there exists, for $q \in \{1, \ldots, Q\}$, $\mathbf{x}^q \in I^q(\mathbf{y}^q)$. It follows that: $\sum_{q=1}^{Q} \mathbf{w}^{q\prime}\mathbf{x}^q \geq \sum_{q=1}^{Q} C^q(\mathbf{y}^q, \mathbf{w}^q)$. As $\mathbf{x}$ (and thus $\mathbf{x}^q$, $\forall q$) is arbitrary, we can choose the cost minimizer levels, which gives:

$$C(\mathbf{y}, \mathbf{w}) \geq \sum_{q=1}^{Q} C^q(\mathbf{y}^q, \mathbf{w}^q). \tag{75}$$

Combining the two previous equations end the proof.