Opening the "Black Box" of Efficiency Measurement: Input Allocation in Multi-Output Settings*

Laurens Cherchye[†], Bram De Rock[‡], Bart Dierynck[§], Filip Roodhooft[¶] and Jeroen Sabbe[‖]

Abstract

We develop a new Data Envelopment Analysis (DEA)-based methodology for measuring the efficiency of Decision Making Units (DMUs) characterized by multiple inputs and multiple outputs. The distinguishing feature of our method is that it explicitly includes information about output-specific inputs and joint inputs in the efficiency evaluation. This method contributes to opening the "black box" of efficiency measurement in two different ways. First, including information on the input allocation substantially increases the discriminatory power of the efficiency measurement. Second, it allows us to decompose the efficiency value of a DMU into output-specific efficiency values which facilitates the identification of the outputs the manager should focus on to remedy the observed inefficiency. We demonstrate the usefulness and managerial implications of our methodology by means of a unique data set collected from the Activity Based Costing (ABC) system of a large service company with 290 DMUs.

Keywords: efficiency measurement, DEA, input allocation, efficiency decomposition, ABC systems

---

*We thank the Area Editor Dan Adelman, an Associate Editor and two anonymous referees for helpful comments, which substantially improved the paper. We also gratefully acknowledge the helpful comments of Jasmijn Bol, Thomas Demuynck, Eva Labro and participants of the 2010 New Directions in Management Accounting Conference on an earlier version of this paper.

†Center for Economic Studies, University of Leuven, and CentER, Tilburg University. E. Sabbelaan 53, B-8500 Kortrijk, Belgium. E-mail: laurens.cherchye@kuleuven.be. Laurens Cherchye gratefully acknowledges financial support from the Research Fund KULeuven through the grant STRT1/08/004.

‡ECARES-ECORE, Université Libre de Bruxelles. Avenue F.D. Roosevelt 50, CP 114, B-1050 Brussels, Belgium. E-mail: bderock@ulb.ac.be. Bram De Rock gratefully acknowledges the European Research Council (ERC) for his starting grant.

§Department of Accounting, Tilburg School of Economics and Management, Tilburg University. PO Box 90153, 5000 LE Tilburg, The Netherlands. Email: b.dierynck@uvt.nl. Bart Dierynck gratefully acknowledges financial support from the Research Foundation Flanders and the Limperg Institute.

¶Accounting, Finance and Insurance, Katholieke Universiteit Leuven and Accounting & Finance, Vlerick Leuven Gent Management School. Naamsestraat 69, B-3000 Leuven, Belgium. E-mail: filip.roodhooft@econ.kuleuven.be

‖Center for Economic Studies, Katholieke Universiteit Leuven. E. Sabbelaan 53, B-8500 Kortrijk, Belgium. E-mail: jeroen.sabbe@kuleuven-kortrijk.be . Jeroen Sabbe gratefully acknowledges Financial support from the Research Foundation Flanders.

1 Introduction

Efficiency analysis of production activities is an important issue for practitioners as well as an area of contemporary interest in both the operations research and economics literature (see, for example, Färe, Grosskopf and Lovell (1994), Cooper, Seiford and Tone (2000), Fried, Lovell and Schmidt (2008), and Cook and Seiford (2009) for reviews). The goal of such analysis is to evaluate the efficiency of a DMU (i.e. Decision Making Unit, which is typically a business unit, office or branch of a private or public sector company) by comparing its input-output performance to that of other DMUs operating in a similar technological environment (typically other business units, offices or branches of the same company). Amongst the efficiency measurement techniques, Data Envelopment Analysis (DEA) has become popular both as an analytical research instrument and as a practical decision-support tool. DEA is a production frontier technique with the distinguishing feature that it is nonparametric in nature, which means that it does not resort to some (typically unverifiable) parametric/functional specifications for the production technology but rather "lets the data speak for themselves".

Still, existing DEA methods essentially provide a "black box" treatment of efficient production behavior, because they only use information on inputs and outputs (and sometimes their prices) to evaluate the efficiency of each DMU. What happens inside the "black box", i.e. how inputs and outputs are exactly linked to each other, does not enter the analysis. However, including such information can improve the discriminatory power of efficiency models without needing to resort to unverifiable assumptions. In this study, we develop a DEA-based methodology for efficiency analysis that explicitly includes information about the allocation of inputs to outputs. In the application, we use Activity Based Costing (ABC) data of a large service company with 290 DMUs to show the practical relevance and managerial implications of our newly developed methodology.

The methodology we develop is rooted in the structural efficiency measurement approach initiated by Afriat (1972), Hanoch and Rothschild (1972), Diewert and Parkan (1983) and Varian (1984).[1] This approach starts from a structural model of efficient production behavior and characterizes inefficiency as deviations from this model. Cherchye, De Rock and Vermeulen (2008) adapted this approach to a multi-output setting that specifically accounts for economies of scope in production. The distinguishing feature of their methodology is that it explicitly recognizes that each different output is characterized by its own production technology, while accounting for interdependencies between the different output-specific technologies. Building on the original idea of Cherchye, De Rock and Vermeulen (2008), we propose an efficiency measurement method that distinguishes between output-specific inputs and joint inputs.[2] The unique feature of our methodology is that we explicitly include information about the allocation of the output-specific inputs to the outputs. This practice opens

[1]See also Banker and Maindiratta (1988) for an early study on the interrelationship between DEA and this structural approach to analyzing efficient production behavior.

[2]Output-specific inputs are inputs that can be fully allocated to an output. For instance, when the input "labor" is used to produce two products and we can observe that 30% of labor time is used for product 1 and 70% for product 2, then labor can be decomposed into output-specific inputs "labor product 1" and "labor product 2". By contrast, joint inputs cannot be allocated to specific outputs. A typical example of a joint input is the compensation package of a CEO.

the black box of efficiency measurement in two different ways. First, including information on the allocation of output-specific inputs substantially increases the discriminatory power of the efficiency measurement: our efficiency measurement method has more power to identify inefficient production behavior. In turn, this should lead to more actions for efficiency improvement and, consequently, higher realized cost reductions. Second, our methodology allows us to decompose the overall efficiency score of a DMU into output-specific efficiency scores and their respective weights in the DMU's overall efficiency. Such a decomposition is particularly attractive from a practical point of view, as it directly identifies the outputs on which DMU managers should principally focus to remedy the observed inefficiency. Thus, our methodology should lead to more improvement actions and support managers to focus these improvement actions on the sources that contribute the most to the observed inefficiency.

As we describe in detail in the following sections, the benefits of our methodology hinge on the availability of information about the allocation of inputs to outputs. Although perfect information about the allocation of inputs to outputs is hardly ever available, many large companies – which are typically considered in efficiency analyses – have well-developed costing systems that provide information about the allocation of inputs (i.e. costs) to outputs (i.e. products). Such information is often used for supporting various strategic and operational decisions such as pricing, product development, and product mix decisions, but can also be used to define and allocate output-specific inputs, which is a core element of our methodology.[3] In this study, we will demonstrate the usefulness of our methodology by means of an empirical application that uses data coming from an ABC system (see Cooper and Kaplan (1988)). The distinguishing feature of ABC is that costs (or inputs) are first allocated to activities (i.e. the first stage of the ABC system) and, subsequently, these activity costs are allocated to the products (or outputs). Compared to other costing methodologies, which often allocate costs to products based on the produced quantities of the different products, ABC gives a much clearer and more accurate picture of the economics of the operations. The use of data from an ABC system for efficiency assessments has, to our knowledge, not been documented before and offers rich ground for practical improvements as well as for research on the interface between operations research and accounting.

The remainder of the paper is organized as follows. Section 2 introduces our efficiency evaluation methodology. In particular, we show how to use information on output-specific and joint inputs in cost efficiency analysis. This will provide a primal interpretation of our efficiency evaluation method. In Section 3, we then introduce a dual interpretation of the same methodology, which uses the same input information in the context of technical efficiency analysis. Section 4 presents our empirical application and discusses the managerial implications. Here, we also demonstrate the usefulness of ABC systems when using our methodology in practice. Section 5 concludes and presents some opportunities for future research.

[3] Note that the costing literature uses "expenses" or "costs" rather than "inputs". However, expenses, costs, and inputs all refer to "resources that are consumed to produce outputs".

2 Primal orientation: cost efficiency

This section sets out our approach to analyze cost efficiency in multi-output settings. After introducing some necessary notation and terminology, we will introduce our multi-output efficiency criterion. In turn, this will allow us to define our cost efficiency measure. Essentially, the cost efficiency orientation followed here complies with the "multiplier" interpretation of DEA models. Specifically, in DEA terminology, multipliers then refer to so-called "shadow prices" for defining DMU efficiency. At the end of Section 2.4, we will provide such a shadow price (or multiplier) formulation of our multi-output cost efficiency measure. In Section 3, we will present a dual interpretation of the same cost efficiency measure. Specifically, we will show that our cost efficiency measure is dually equivalent to a measure of technical input efficiency that is specially defined for multi-output settings. As we will explain, this dual representation will comply with the "envelopment" formulation of DEA models.

2.1 Preliminaries

Practical efficiency analysis starts from a data set with T DMUs, which produce M outputs. As indicated above, at the input side, we make the distinction between output-specific inputs and joint inputs. Specifically, we assume N^{spec} output-specific inputs (that can be allocated to particular outputs) and N^{join} joint inputs (that cannot be allocated). In what follows, we will assume the allocation of the output-specific inputs is observed; this will effectively be the case for our empirical application in Section 4 (which will use an ABC system for the input allocation). At this point, however, it is worth to add that our method can also be applied if output-specific inputs are not observed. We will discuss the corresponding extension in Section 2.5. Next, in this section we will first assume that the data set also contains the prices of the (output-specific and joint) inputs. We will relax this assumption later on (in Section 2.4). In fact, exact price information will not be available for our empirical application in Section 4.

More formally, we use the following notation for the observed quantities and prices of each DMU t $(1 \leq t \leq T)$. First, we observe an M-vector of outputs $\mathbf{y}_t \in \mathbb{R}_+^M$; we use $\mathbf{y}_t = \left(y_t^1, ..., y_t^M\right)$ with each entry y_t^m representing the amount that DMU t produces of the m-th output $(1 \leq m \leq M)$. Next, we observe an N^{spec}-vector of output-specific inputs $\mathbf{q}_t^m \in \mathbb{R}_+^{Nspec}$ for each individual output m, and an N^{join}-vector of joint inputs $\mathbf{Q}_t \in \mathbb{R}_+^{Njoin}$. Correspondingly, we observe a price vector $\mathbf{p}_t \in \mathbb{R}_+^{Nspec}$ for the output-specific inputs and a price vector $\mathbf{P}_t \in \mathbb{R}_+^{Njoin}$ for the joint inputs. The full data set can be summarized as

$$S = \left\{ \left(\mathbf{y}_t, \mathbf{q}_t^1, ..., \mathbf{q}_t^M, \mathbf{Q}_t, \mathbf{p}_t, \mathbf{P}_t\right) | t = 1, ..., T \right\}.$$

In what follows, we use $\mathbf{p}_t' \left(\sum_{m=1}^M \mathbf{q}_t^m\right) + \mathbf{P}_t'\mathbf{Q}_t = z_t$ where z_t is the budget (or cost) associated with DMU t.[4]

As discussed above, a specific feature of our framework here is that it explicitly recognizes that each different output is characterized by its own production technology. At the same

[4] Throughout, we use the symbol $'$ to indicate transposition of a matrix, i.e. $\mathbf{A}'$ represents the transpose of matrix $\mathbf{A}$.

time, we account for interdependencies between the different output-specific technologies through the joint inputs $\mathbf{Q}$; as discussed in the introduction, including joint inputs allows for economies of scope in production (see Cherchye, De Rock and Vermeulen, 2008). To formalize this idea, we characterize the production technology of each output m by input requirement sets $I^m(y^m)$ $(1 \leq m \leq M)$, which contain all the combinations of output-specific and joint inputs $(\mathbf{q}^m, \mathbf{Q})$ that can produce the output quantity y^m:

$$I^m(y^m) = \{(\mathbf{q}^m, \mathbf{Q}) \in \mathbb{R}_+^{Nspec} \times \mathbb{R}_+^{Njoin} | (\mathbf{q}^m, \mathbf{Q}) \text{ can produce } y^m\}.$$

In what follows, we will assume that the production technology satisfies the axiom of nested input requirement sets.

Axiom 1 (nested input sets) : $y^m \geq y^{m\star} \Rightarrow I^m(y^m) \subseteq I^m(y^{m\star})$.

This says that if a particular input combination $(\mathbf{q}^m, \mathbf{Q})$ can produce the output quantity y^m, then it can also produce any lower quantity $y^{m\star}$. Essentially, this axiom of nested input requirement sets implies that outputs are freely disposable. Free output disposability is a standard assumption in the DEA literature. See, for example, Varian (1984) and Tulkens (1993) for discussion.[5]

To assess multi-output cost efficiency, we will need a way to evaluate the joint inputs. Here, we use the concept of implicit prices.

Definition 1 (implicit prices) : *For DMU t, with prices $\mathbf{P}_t$ for the joint inputs, implicit prices are any vectors $\mathbf{P}_t^m \in \mathbb{R}_+^{Njoin} (m = 1, \ldots, M)$ that satisfy $\sum_{m=1}^{M} \mathbf{P}_t^m = \mathbf{P}_t$.*

In words, implicit prices represent the fraction of the aggregate prices of the joint inputs that are borne by the different outputs. This is an intuitive concept from a costing perspective, where some overhead costs are sometimes used by multiple outputs (i.e. they represent joint inputs), but it is unknown to the cost accountant or empirical analyst at which ratio this happens. In theory it should be possible to identify the implicit prices if one knew the exact "contribution" of each individual output to the joint input. However, in practice this information is usually not available to the empirical analyst, and therefore we define as implicit prices any price vectors that sum to the observed prices. In Section 2.5, we will indicate the possible extension of our method to include information on (feasible ranges of) implicit prices if such information would be available.

At this point, it is worth to stress that the concept of implicit input prices also has a solid theoretical foundation. Specifically, it parallels the notions of Lindahl prices associated with Pareto efficient public goods provision. In particular, in this context Lindahl prices represent the willingness-to-pay of the different consumers of the public good. Pareto efficiency then requires that these Lindahl prices (summed over the different consumers) must exactly equal the market price of the public good (see, for example, Myles (1995)). In our case, joint inputs $\mathbf{Q}$ clearly have a "public good" nature as they simultaneously benefit the production

[5] Actually, the assumption of nested input requirement sets implies that the cost minimization criterion in Definition 2 does have empirical content, i.e. production data need not automatically satisfy the criterion. This follows from a ready adaptation of Varian's (1984) argument that applies to the case of single output production.

of the different outputs m. Given this, the condition that the output-specific implicit prices $\mathbf{P}_t^m$ must exactly add up to the observed prices $\mathbf{P}$ directly complies with the Lindahl pricing condition for efficient public goods provision. We refer to Cherchye, De Rock and Vermeulen (2008) and Cherchye, Demuynck, De Rock and De Witte (2011) for a formal discussion on this relation between implicit prices as used here and Lindahl prices for public goods.[6]

2.2 Cost efficiency criterion

The basic idea underlying our empirical criterion for multi-output cost efficiency is that, in practice, we do not observe the true input requirement sets or implicit prices for the joint inputs. For some observed DMU t, the criterion checks whether we can specify the input sets (consistent with Axiom 1) and implicit prices under which the observed inputs of DMU t can be labelled as cost efficient. In formal terms, this obtains the next definition.

Definition 2 (cost efficiency) : *DMU t is multi-output cost efficient if and only if there exist for each output m $(1 \leq m \leq M$ $)$ input requirement sets $I^m(y_t^m)$ that satisfy Axiom 1 and implicit prices $\mathbf{P}_t^m \in \mathbb{R}_+^{Njoin}$ such that, for each output m,*

(CE-1) *we have $(\mathbf{q}_t^m, \mathbf{Q}_t) \in I^m(y_t^m)$, and*

(CE-2) *it holds that $\mathbf{p}_t^{'}\mathbf{q}_t^m + (\mathbf{P}_t^m)^{'}\mathbf{Q}_t = \min_{(\mathbf{q}^m,\mathbf{Q})\in I^m(y_t^m)} \mathbf{p}_t^{'}\mathbf{q}^m + (\mathbf{P}_t^m)^{'}\mathbf{Q}$.*

Importantly, this cost efficiency criterion is intrinsically nonparametric. It does not need a prior specification of the input sets $I^m(y_s^m)$. It only requires that there exists *at least* one specification of these sets that is consistent with Axiom 1 and simultaneously meets the conditions (CE-1) and (CE-2).

Condition (CE-1) is essentially a feasibility requirement: it states that, for DMU t, the input requirement sets $I^m(y_t^m)$ must be such that the inputs $\mathbf{q}_t^m$ and $\mathbf{Q}_t$ can effectively produce the output y_t^m (i.e. $(\mathbf{q}_t^m, \mathbf{Q}_t) \in I^m(y_t^m)$). As such, the condition guarantees that we consider input requirement sets such that the observed input of DMU t effectively can produce the observed output of this DMU. In other words, it makes sure that we consider technologies for which the input-output combinations that we observe are certainly feasible (compare with Axiom 4 below).

Next, condition (CE-2) imposes multi-output cost efficiency: for the given set $I^m(y_t^m)$ and implicit prices $\mathbf{P}_t^m$, DMU t must produce each output m at a minimal cost. We say that DMU t is cost efficient if this cost minimization condition is met. Alternatively, if (CE-2) is not satisfied, i.e.

$$\mathbf{p}_t^{'}\mathbf{q}_t^m + (\mathbf{P}_t^m)^{'}\mathbf{Q}_t > \min_{(\mathbf{q}^m,\mathbf{Q})\in I^m(y_t^m)} \mathbf{p}_t^{'}\mathbf{q}^m + (\mathbf{P}_t^m)^{'}\mathbf{Q},$$

then we conclude cost inefficient behavior. In such a situation, the same output vector could have been produced at a lower cost *for any feasible* specification of the implicit prices $\mathbf{P}_t^m$

[6]Here, it is also useful to refer to related discussions in Chiappori (1988) and Cherchye, De Rock and Vermeulen (2007, 2011), who present a nonparametric methodology for analyzing (Pareto efficient) multi-person consumption that is formally close to our methodology for analyzing multi-output production. In their consumption setting, these authors equally evaluate goods that are jointly consumed in terms of implicit prices that can be interpreted as Lindahl prices.

and input requirement sets $I^m(y_t^m)$. For any feasible prices $\mathbf{P}_t^m$ and sets $I^m(y_t^m)$, there is at least one output m that is not produced at a minimal cost. In this respect, in Section 2.3 we discuss the possibility to decompose multi-output cost (in)efficiency into output-specific cost (in)efficiencies.

Definition 2 is not directly applicable in practice. For a given set of observations S, there are typically infinitely many specifications of the sets $I^m(y_t^m)$ that meet condition (CE-1) and infinitely many specifications of the implicit prices $\mathbf{P}_t^m$ that satisfy the adding up condition $\sum_{m=1}^M \mathbf{P}_t^m = \mathbf{P}_t$ (in Definition 1). In principle, checking the cost efficiency criterion in Definition 2 would require us to verify condition (CE-2) for each possible specification of $I^m(y_t^m)$ and $\mathbf{P}_t^m$. Clearly, this is not feasible in finite time.

Fortunately, we can derive an equivalent formulation of the multi-output cost efficiency criterion that can be used in practice, because it can be verified through linear programming. It is given in the next result.

Proposition 1 : *DMU t is multi-output cost efficient if and only if there exist implicit prices* $\mathbf{P}_t^m \in \mathbb{R}_+^{Njoin}$ $(1 \leq m \leq M)$ *such that, for each output m, it holds that: if, for some DMU s,* $y_s^m \geq y_t^m$, *then* $\mathbf{p}_t^{'}\mathbf{q}_t^m + (\mathbf{P}_t^m)^{'}\mathbf{Q}_t \leq \mathbf{p}_t^{'}\mathbf{q}_s^m + (\mathbf{P}_t^m)^{'}\mathbf{Q}_s$.

Thus, DMU t is multi-output cost efficient if and only if it produces every output m at the minimal cost defined over the set of observed DMUs, while using implicit prices to evaluate the joint inputs. In view of our following exposition, it is useful to reformulate this cost minimization criterion for each output m as follows:

$$\mathbf{p}_t^{'}\mathbf{q}_t^m + (\mathbf{P}_t^m)^{'}\mathbf{Q}_t = \min_{s \in D_t^m} \left(\mathbf{p}_t^{'}\mathbf{q}_s^m + (\mathbf{P}_t^m)^{'}\mathbf{Q}_s \right), \tag{1}$$

where we use

$$D_t^m = \{s | y_t^m \leq y_s^m\}, \tag{2}$$

i.e. the set D_t^m captures all DMUs s that produce at least the output y_t^m $(y_t^m \leq y_s^m)$. In the next section, we show that this efficiency criterion implies a natural efficiency measure. Subsequently, we show in Section 2.4 that this efficiency measure can be computed through linear programming, which makes it easily implementable.

2.3 Cost efficiency measurement

Suppose we want to evaluate DMU t in terms of the multi-output cost efficiency criterion in Definition 2. We start from the cost minimization condition (1) for each output m. For a given specification of the implicit prices $\mathbf{P}_t^m$, we can define the minimal cost for output m as

$$c_t^m(\mathbf{P}_t^m) = \min_{s \in D_t^m} \left(\mathbf{p}_t^{'}\mathbf{q}_s^m + (\mathbf{P}_t^m)^{'}\mathbf{Q}_s \right), \tag{3}$$

so that condition (1) requires $\mathbf{p}_t^{'}\mathbf{q}_t^m + (\mathbf{P}_t^m)^{'}\mathbf{Q}_t = c_t^m(\mathbf{P}_t^m)$. When considering all outputs m together, this naturally suggests the following measure of cost efficiency:

$$CE_t\left(\mathbf{P}_t^1, ..., \mathbf{P}_t^M\right) = \frac{\sum_{m=1}^M c_t^m(\mathbf{P}_t^m)}{\sum_{m=1}^M (\mathbf{p}_t^{'}\mathbf{q}_t^m + (\mathbf{P}_t^m)^{'}\mathbf{Q}_t)} = \frac{\sum_{m=1}^M c_t^m(\mathbf{P}_t^m)}{\sum_{m=1}^M \mathbf{p}_t^{'}\mathbf{q}_t^m + \mathbf{P}_t^{'}\mathbf{Q}_t}.$$

Clearly, we have $0 \leq CE_t\left(\mathbf{P}_t^1, ..., \mathbf{P}_t^M\right) \leq 1$, with lower values indicating less cost efficiency (or more cost inefficiency). The value of $CE_t\left(\mathbf{P}_t^1, ..., \mathbf{P}_t^M\right)$ has a convenient degree interpretation: for given $\mathbf{P}_t^m$, it captures the extent to which the actual cost $\left(\sum_{m=1}^M \mathbf{p}_t'\mathbf{q}_t^m + \mathbf{P}_t'\mathbf{Q}_t\right)$ exceeds the minimal cost $\left(\sum_{m=1}^M c_t^m\right)$ for the (multi-dimensional) output that is produced.

However, the cost efficiency measure $CE_t\left(\mathbf{P}_t^1, ..., \mathbf{P}_t^M\right)$ is not directly useful because it requires a prior specification of the implicit prices $\mathbf{P}_t^m$. In empirical applications, we typically do not observe these prices. In this respect, we recall that Definition 2 (only) requires that there exists *at least one* specification of the implicit prices such that each observation is cost efficient. As such, we can use the following cost efficiency measure in practical efficiency analysis:

$$CE_t = \max_{\mathbf{P}_t^m \in \mathbb{R}_+^{Njoin} : \sum_m \mathbf{P}_t^m = \mathbf{P}_t} CE_t\left(\mathbf{P}_t^1, ..., \mathbf{P}_t^M\right). \tag{4}$$

In words, this measure selects those implicit prices that maximize the cost efficiency of DMU t. Intuitively, these implicit prices can be interpreted as most favorable prices for evaluating the joint inputs. In fact, such most favorable pricing is implicitly used in DEA; see our discussion of LP-2 below.

Similar to before, we have that $0 \leq CE_t \leq 1$, with lower values indicating less cost efficiency; and the degree interpretation of $CE_t\left(\mathbf{P}_t^1, ..., \mathbf{P}_t^M\right)$ carries over to CE_t (but now for the endogenously selected $\mathbf{P}_t^m$). Clearly, DMU t meets the multi-output cost efficiency criterion in Definition 2 if and only if $CE_t = 1$.

Importantly, the multi-output cost efficiency measure CE_t can naturally be decomposed in terms of output-specific cost efficiencies. To see the decomposition, let $\mathbf{P}_t^{m*}$ solve the max problem in (4), i.e.

$$\left(\mathbf{P}_t^{1*}, ..., \mathbf{P}_t^{M*}\right) = \arg \max_{\mathbf{P}_t^m \in \mathbb{R}_+^{Njoin} : \sum_m \mathbf{P}_t^m = \mathbf{P}_t} CE_t\left(\mathbf{P}_t^1, ..., \mathbf{P}_t^M\right). \tag{5}$$

Correspondingly, we have

$$CE_t = \frac{\sum_{m=1}^M c_t^m\left(\mathbf{P}_t^{m*}\right)}{\sum_{m=1}^M \mathbf{p}_t'\mathbf{q}_t^m + \mathbf{P}_t'\mathbf{Q}_t}.$$

Using this, we can write

$$CE_t = \sum_{m=1}^M w^m CE_t^m, \tag{6}$$

where

$$w_t^m = \frac{\mathbf{p}_t'\mathbf{q}_t^m + \left(\mathbf{P}_t^{m*}\right)'\mathbf{Q}_t}{\sum_{m=1}^M \mathbf{p}_t'\mathbf{q}_t^m + \mathbf{P}_t'\mathbf{Q}_t} \text{ and } CE_t^m = \frac{c_t^m\left(\mathbf{P}_t^{m*}\right)}{\mathbf{p}_t'\mathbf{q}_t^m + \left(\mathbf{P}_t^{m*}\right)'\mathbf{Q}_t}.$$

In this decomposition, CE_t^m measures the cost efficiency of DMU t in producing output m, while w_t^m represents the weight of this output in the overall (multi-output) cost efficiency measure CE_t. More specifically, the output-specific efficiency measure CE_t^m (always between 0 and 1) expresses how cost efficient DMU t is at producing output m. Next, the weight w_t^m (also between 0 and 1) represents the share of the total budget that is allocated to output m

(for the given implicit prices $\mathbf{P}_t^{m*}$). Ex post, this can be interpreted as the weight allocated to output m in the calculation of the multi-output efficiency measure CE_t.

We believe the decomposition in (6) has substantial practical value because the output-specific efficiency measures can guide DMUs when evaluating the cause of their observed inefficiency as well as when planning actions to improve efficiency. In Section 4, we will illustrate the application of the decomposition for managerial purposes.

2.4 Practical implementation

A particularly attractive feature of the measure CE_t in (4) is that it can be computed through linear programming (LP). Actually, the solution of the LP problem also gives the implicit prices $\mathbf{P}_t^{m*}$ that solve the maximization problem in equation (5). In turn, this enables us to compute the output-specific cost efficiencies CE_t^m and the corresponding weights w_t^m, and so to conduct the decomposition of CE_t in equation (6).

As we will explain, the maximization problem in equation (4) is equivalent to the following LP problem (**LP-1**):

$$
\begin{aligned}
& CE_t = \max_{c_t^m \geq 0, \mathbf{P}_t^m \in \mathbb{R}_+^{Njoin}} \frac{\sum_{m=1}^M c_t^m}{\sum_{m=1}^M \mathbf{p}_t^{\prime} \mathbf{q}_t^m + \mathbf{P}_t^{\prime} \mathbf{Q}_t} \\
& s.t. \\
& \text{(C-1)} \ \sum_{m=1}^M \mathbf{P}_t^m = \mathbf{P}_t \\
& \text{(C-2)} \ \forall m: c_t^m \leq \mathbf{p}_t^{\prime} \mathbf{q}_s^m + \left(\mathbf{P}_t^m\right)^{\prime} \mathbf{Q}_s \ \forall s \in D_t^m
\end{aligned}
$$

In this problem, the constraints $\mathbf{P}_t^m \in \mathbb{R}_+^{Njoin}$ and (C-1) make sure that the endogenously selected implicit prices $\mathbf{P}_t^m$ satisfy Definition 1. Next, for given prices $\mathbf{P}_t^m$, the constraint (C-2) ensures that c_t^m in LP-1 satisfies equation (3), which defines $c_t^m\left(\mathbf{P}_t^m\right)$. As a result, we obtain that the solution to LP-1 effectively solves the maximization problem in equation (4) and vice versa, i.e. the values $\mathbf{P}_t^m$ defined in equation (4) and the corresponding values $c_t^m\left(\mathbf{P}_t^m\right)$ solve LP-1.

So far, we have assumed that the input price vectors $\mathbf{p}_t$ and $\mathbf{P}_t$ are exactly observed. However, in many empirical applications such exact price information is not available.[7] In such a situation, we use the data set

$$
S = \left\{ \left(\mathbf{y}_t, \mathbf{q}_t^1, ..., \mathbf{q}_t^M, \mathbf{Q}_t \right) | t = 1, ..., T \right\}.
$$

Attractively, our cost efficiency measures (including the corresponding LP characterization) can easily be adjusted to account for such incomplete price information. Consistent with usual practice in DEA, we use "most favorable" prices for evaluating the output-specific and joint inputs in the absence of exact price information: we adjust LP-1 so that it selects prices that maximize the efficiency of DMU t. In a certain sense, such most favorable prices may be interpreted as *shadow prices* that support cost efficient behavior of the evaluated DMU. Intuitively, (most favorable) shadow prices give each DMU t the "benefit of the doubt" in the efficiency evaluation exercise. In DEA terminology, shadow prices are referred to as

[7] See, for example, Kuosmanen, Cherchye and Sipiläinen (2006) for a discussion of instances where reliable price information is not readily available. Our application in Section 4 contains another example.

"multipliers" (see, for example, Cooper, Seiford and Tone (2000))[8]. As such, our method for cost efficiency evaluation (under shadow prices) corresponds to the multiplier formulation of standard DEA models.

More formally, the use of shadow prices $\mathbf{p}_t$ and $\mathbf{P}_t$ for the inputs obtains the following LP problem (**LP-2**):

$$
\begin{aligned}
&\widehat{CE}_t = \max_{\substack{c_t^m \geq 0, \mathbf{P}_t^m \in \mathbb{R}_+^{Njoin},\\ \mathbf{P}_t \in \mathbb{R}_+^{Njoin}, \mathbf{p}_t \in \mathbb{R}_+^{Nspec}}} \sum_{m=1}^{M} c_t^m \\
&s.t. \\
&\text{(C-1) } \sum_{m=1}^{M} \mathbf{P}_t^m = \mathbf{P}_t \\
&\text{(C-2) } \forall m : c_t^m \leq \mathbf{p}_t^{\prime} \mathbf{q}_s^m + \left(\mathbf{P}_t^m\right)^{\prime} \mathbf{Q}_s \ \forall s \in D_t^m \\
&\text{(C-3) } \sum_{m=1}^{M} \mathbf{p}_t^{\prime} \mathbf{q}_t^m + \mathbf{P}_t^{\prime} \mathbf{Q}_t = 1
\end{aligned}
$$

This LP problem has a readily similar interpretation as LP-1. The only difference is the normalization constraint (C-3) in LP-2. This constraint implies that we can give the objective function of LP-2 a similar ratio interpretation as the objective function of LP-1: because of (C-3) we have $\sum_{m=1}^{M} c_t^m = \left(\sum_{m=1}^{M} c_t^m / \sum_{m=1}^{M} \mathbf{p}_t^{\prime} \mathbf{q}_t^m + \mathbf{P}_t^{\prime} \mathbf{Q}_t\right)$.[9]

2.5 Extensions

As explained above, the shadow price problem LP-2 complies with the multiplier formulation of DEA models. Given this, we can consider alternative extensions that include existing DEA tools in our framework, to deal with specific issues that can be relevant in practical applications. Here, we focus on two such extensions that are particularly interesting for our (multi-output) cost efficiency setting.

A first extension pertains to the fact that an efficiency analysis based on LP-2 can be strengthened by imposing price information, which then takes the form of additional constraints that define a feasible range for the relative prices. For example, such shadow price (or multiplier) constraints may rule out the extreme cases where the relative price of a commodity approaches zero or infinity. Actually, given our specific multi-output setting, such price restrictions may also pertain to the implicit prices that are used for valuing the joint inputs in the cost efficiency evaluation. The technical questions related to incorporating shadow price restrictions have been discussed extensively in a DEA context, most commonly under the label "weight restrictions" or "assurance regions" (see, for example, Allen, Athanassopoulos, Dyson and Thanassoulis (1997) and Pedraja-Chaparro, Salina-Jimenez and Smith (1997), for surveys, and Kuosmanen, Cherchye and Sipiläinen (2006) for more recent developments). These tools are readily adapted to the current set-up. Typically, DEA shadow price restrictions are linear and, as such, they do not interfere with the linear nature of LP-2. As a specific illustration, we will use price restrictions in our application in Section 4.

[8] Cherchye, Moesen, Rogge and Van Puyenbroeck (2007) provide a detailed discussion of the "benefit of the doubt"-interpretation of DEA models in the specific context of composite indicator construction.

[9] Charnes and Cooper (1962) originally proposed to use a normalization constraint to convert a fractional programming problem into an equivalent linear programming problem. In fact, Charnes, Cooper and Rhodes (1978) also used this normalization procedure in their first DEA paper.

A second extension concerns the possibility that, in practice, we do not observe the output-specific input vectors $\mathbf{q}_s^m$ but only the aggregate vector $\mathbf{q}_s = \sum_{m=1}^{M} \mathbf{q}_t^m$. In such a case, we can use a procedure outlined by Cook, Habadou and Teunter (2000) and Cook and Habadou (2001) in a closely similar DEA context.[10] Essentially, this procedure deals with unobserved output-specific input vectors under the assumption that the fraction of the input vector $\mathbf{q}_t$ allocated to each different output m is the same for all DMUs t.[11] Formally, we let the vector $\boldsymbol{\alpha}^m = (\alpha^{m,1}, \ldots, \alpha^{m,Nspec}) \in \mathbb{R}_+^{Nspec}$ represent the fractions of the different entries of $\mathbf{q}_t$ that are allocated to output m. Then, we get

$$\text{(C-4)} \sum_{m=1}^{M} \boldsymbol{\alpha}^m = \mathbf{1},$$

where $\mathbf{1}$ is an $Nspec$-dimensional vector containing only ones, and

$$\forall t : \mathbf{q}_t^m = \boldsymbol{\alpha}^m \odot \mathbf{q}_t,$$

with $\odot$ the Hadamard or element-by-element product. In terms of problems LP-1 and LP-2, this means adding the decision variables $\boldsymbol{\alpha}^m \in \mathbb{R}_+^{Nspec}$ and constraint (C-4). In addition, we must replace constraint (C-2) by

$$\text{(C-5)}\ \forall m : c_t^m \leq \mathbf{p}_t^{'} (\boldsymbol{\alpha}^m \odot \mathbf{q}_t) + (\mathbf{P}_t^m)^{'} \mathbf{Q}_s \ \forall s \in D_t^m.$$

For the problem (LP-1), with input price vectors $\mathbf{p}_t$ known to the empirical analyst, adding the constraint (C-4) and replacing (C-2) by (C-5) obviously does not interfere with its linear nature. As such, this procedure for dealing with unobserved output-specific input vectors is directly implementable through linear programming.

However, for the problem (LP-2), which uses shadow prices because the vector $\mathbf{p}_t$ is unobserved, the constraint (C-5) is nonlinear in the variables $\mathbf{p}_t$ and $\boldsymbol{\alpha}^m$. Following Cook, Habadou and Teunter, we can remedy this problem by first noting that (C-5) is equivalent to

$$\forall m : c_t^m \leq (\mathbf{p}_t \odot \boldsymbol{\alpha}^m)^{'} \mathbf{q}_t + (\mathbf{P}_t^m)^{'} \mathbf{Q}_s \ \forall s \in D_t^m,$$

and subsequently replacing the (Hadamard) product of the variables $\mathbf{p}_t$ and $\boldsymbol{\alpha}^m$ by a new variable $\boldsymbol{\pi}_t \in \mathbb{R}_+^{Nspec}$, i.e. $\boldsymbol{\pi}_t = \mathbf{p}_t \odot \boldsymbol{\alpha}^m$. Thus, for LP-2 we do not use (C-5) but, instead, the new constraint

$$\text{(C-6)}\ \forall m : c_t^m \leq \boldsymbol{\pi}_t^{'} \mathbf{q}_t + (\mathbf{P}_t^m)^{'} \mathbf{Q}_s \ \forall s \in D_t^m,$$

which is clearly linear. At this point, the constraint (C-4) may seem redundant in combination with (C-6), as the variables $\boldsymbol{\alpha}^m$ no longer appear in (C-6). This is true (only) if no shadow price (or multiplier) constraint applies to the vector $\mathbf{p}_t$ (except from non-negativity). In such a case, replacing $\mathbf{p}_t \odot \boldsymbol{\alpha}^m$ by $\boldsymbol{\pi}_t$ does not put any restriction on the values $\boldsymbol{\pi}_t$ can

[10] See also Beasley (1995) and Cook and Green (2004) for related ideas.

[11] We remark that the assumption that all DMUs allocate the same fractions of the input $\mathbf{q}$ allocated to the outputs m may be a strong one in practical applications. For example, inspection of the data used in our own application (which contains information on the vectors $\mathbf{q}_t^m$; see Section 4) reveals that the assumption is violated for this particular setting.

take (except from non-negativity).

However, the situation alters if the vector $\mathbf{p}_t$ is subject to shadow price constraints. In that case, the constraint (C-4) effectively turns out to be restrictive. This is extensively discussed by Cook, Habadou and Teunter (2000), who also show how to translate linear constraints on $\mathbf{p}_t$ into linear constraints on $\boldsymbol{\pi}_t$ (while accounting for $\boldsymbol{\pi}_t = \mathbf{p}_t \odot \boldsymbol{\alpha}^m$). These authors' discussion directly carries over to our problem LP-2 with (C-4) added and the original constraint (C-2) replaced by (C-6). For compactness, we do not repeat this here.

3 Dual orientation: technical efficiency

As we indicated, our multi-output cost efficiency measure that makes use of shadow pricing can also be interpreted as a DEA model in multiplier form. Its specific feature as compared to existing DEA measures is its multi-output orientation. Here, it is worth to indicate that DEA models are most often expressed in "envelopment" form. In particular, envelopment DEA models measure the technical efficiency of a DMU with respect to a technical feasibility set (representing the production technology) that envelops the input-output combinations associated with the observed DMUs. In this section, we will introduce a similar technical efficiency measure that is specially tailored for multi-output settings. Subsequently, we will show that this measure is dually equivalent to the cost efficiency measure defined above. This effectively provides the envelopment (technical efficiency) counterpart of the multiplier (cost efficiency) measure presented in Section 2.

3.1 Technical input efficiency

In contrast to cost efficiency, technical efficiency analysis does not use price information, i.e. it starts from a data set

$$S = \left\{ \left(\mathbf{y}_t, \mathbf{q}_t^1, ..., \mathbf{q}_t^M, \mathbf{Q}_t\right) | t = 1, ..., T \right\}.$$

Technical efficiency of a DMU t then quantifies the distance from the output-input combination $\left(\mathbf{y}_t, \mathbf{q}_t^1, ..., \mathbf{q}_t^M, \mathbf{Q}_t\right)$ to the boundary of the technical feasibility set associated with the given production technology. Like before, we define technical feasibilities in terms of input requirement sets $I^m(y^m)$ $(1 \leq m \leq M)$. In our multi-ouput context, an output-input combination $\left(\mathbf{y}, \mathbf{q}^1, ..., \mathbf{q}^M, \mathbf{Q}\right)$ is technically feasible if, for all outputs m, we have that $(\mathbf{q}^m, \mathbf{Q}) \in I^m(y^m)$.

To establish duality with our multi-output cost efficiency measure, we assess technical efficiency in input terms. We use the Debreu-Farrell input efficiency measure, which is the most commonly used efficiency measure in the DEA literature. For DMU t, this measure is defined as[12]

$$TE_t = \min\{\theta | \ \forall m : (\theta \mathbf{q}_t^m, \theta \mathbf{Q}_t) \in I^m(y_t^m)\}. \tag{7}$$

[12] Throughout, we will assume that the production technology has the required properties for the minimum value in (7) to be defined. For example, this will be the case for the empirical version of the measure TE_t (i.e. $\widehat{TE}_t$) that we introduce further on (see (9)).

In words, the measure TE_t captures the maximum equiproportionate reduction (measured by θ) of DMU t's inputs $\left(\mathbf{q}_t^1, ..., \mathbf{q}_t^M, \mathbf{Q}_t\right)$ such that the resulting input vector $\left(= \theta \left(\mathbf{q}_t^1, ..., \mathbf{q}_t^M, \mathbf{Q}_t\right)\right)$ can still produce the output $\mathbf{y}_t$ (i.e. for each output m we have $(\theta \mathbf{q}_t^m, \theta \mathbf{Q}_t) \in I^m(y_t^m)$). Like in our cost efficiency assessment, the specificity of this efficiency measure is that it explicitly recognizes the multi-output nature of the production setting, by considering output-specific input sets $I^m(y_t^m)$ and so accounting for joint inputs as well as output-specific inputs.

In general, for $(\mathbf{q}_t^m, \mathbf{Q}_t) \in I^m(y_t^m)$ we have

$$0 \leq TE_t \leq 1,$$

with $TE_t = 1$ indicating technical input efficiency. For the given technical feasibility set, the measure TE_t defines the maximal (equiproportionate) input reduction that still allows for producing the output $\mathbf{y}_t$. In general, lower values of TE_t indicate greater technical inefficiency.

3.2 Technology axioms

As it is defined in (7), the measure TE_t does not have practical usefulness. It requires knowing the sets $I^m(y_t^m)$. In practice, the empirical analyst typically does not observe the full technical feasibility set but can only use the production information revealed by the observed DMUs. Envelopment DEA models reconstruct the technical feasibility set by starting from these DMUs and additionally using a number of technology axioms. This obtains a technical feasibility set that envelops the observed DMUs.

To obtain a technical efficiency measure that is dually equivalent to the cost efficiency measure introduced in Section 2, we need three technology axioms. Like before, we have to assume that input requirement sets are nested (Axiom 1). The two other axioms are monotonicity and convexity of the sets $I^m(y^m)$. In Section 3.3, we will explain in more detail the specific role of these two technology axioms for establishing a dual relationship between our technical and cost efficiency measures.

Axiom 2 (monotone input sets) : $(\mathbf{q}^m, \mathbf{Q}) \in I^m(y^m)$ *and* $(\mathbf{q}^{m\star}, \mathbf{Q}^\star) \geq (\mathbf{q}^m, \mathbf{Q}) \Rightarrow (\mathbf{q}^{m\star}, \mathbf{Q}^\star) \in I^m(y^m)$.

Axiom 3 (convex input sets) : $(\mathbf{q}^m, \mathbf{Q}) \in I^m(y^m)$ *and* $(\mathbf{q}^{m\star}, \mathbf{Q}^\star) \in I^m(y^m) \Rightarrow \forall \lambda \in [0,1] : \lambda(\mathbf{q}^{m\star}, \mathbf{Q}^\star) + (1-\lambda)(\mathbf{q}^{m\star}, \mathbf{Q}^\star) \in I^m(y^m)$.

Here, monotonicity of $I^m(y^m)$ means that inputs are freely disposable, i.e. more input never reduces the (producible) output. Next, convexity of $I^m(y^m)$ says that, if two input vectors $(\mathbf{q}^m, \mathbf{Q})$ and $(\mathbf{q}^{m\star}, \mathbf{Q}^\star)$ can produce the output y^m, then any convex combination $(= \lambda(\mathbf{q}^{m\star}, \mathbf{Q}^\star) + (1-\lambda)(\mathbf{q}^{m\star}, \mathbf{Q}^\star))$ can also produce the same output. These monotonicity and convexity axioms are often used in DEA analysis. See, for example, Petersen (1990) and Bogetoft (1996) for discussion.

As indicated above, a typical feature of DEA models is that they use a feasibility set that envelops the input-output combinations associated with the observed DMUs. This property is guaranteed by the next axiom:

Axiom 4 (observability means feasibility) : $\left(\mathbf{y}_t, \mathbf{q}_t^1, ..., \mathbf{q}_t^M, \mathbf{Q}_t\right) \in S \Rightarrow \forall m : (\mathbf{q}_t^m,\ \mathbf{Q}_t) \in I^m(y_t^m)$.

Essentially, this axiom says that, if we observe the input $\left(\mathbf{q}_t^1, ..., \mathbf{q}_t^M, \mathbf{Q}_t\right)$ in combination with the output $\mathbf{y}_t$, then this input can certainly produce this output (i.e. $(\mathbf{q}_t^m, \mathbf{Q}_t) \in I^m(y_t^m)$ for all m). Or, what we observe is certainly feasible. We note that this axiom effectively assumes that all input and output data are correctly measured. Obviously, this may be a problem in empirical work. Therefore, in our application (in Section 4) we will also consider an extension of our methodology that accounts for the possibility of measurement errors. As for now, however, we do maintain Axiom 4, mainly to (substantially) simplify the formal exposition.

For a given set of axioms, DEA then provides an empirical construction of the technical feasibility set that satisfies the "minimum extrapolation principle", which means that this construction defines the smallest feasibility set that is consistent with the stated axioms. For our Axioms 1-4, this defines the following empirical estimate of each set $I^m\left(y_t^m\right)$:

$$\widehat{I}^m\left(y_t^m\right) = \left\{ (\mathbf{q}^m, \mathbf{Q}) \left| \begin{array}{c} \sum_{s \in D_t^m} \lambda_s^m \mathbf{q}_s^m \leq \mathbf{q}^m, \sum_{s \in D_t^m} \lambda_s^m \mathbf{Q}_s \leq \mathbf{Q} \\ \sum_{s \in D_t^m} \lambda_s^m = 1, \lambda_s^m \geq 0 \end{array} \right. \right\}, \tag{8}$$

with the set D_t^m defined in (2). In words, the empirical set $\widehat{I}^m\left(y_t^m\right)$ is constructed as the convex-monotone hull of the input vectors $(\mathbf{q}_s^m, \mathbf{Q}_s)$ associated with all DMUs s that produce at least the output y_t^m (i.e. $s \in D_t^m$).

The next result provides a formal statement of the fact that $\widehat{I}^m\left(y_t^m\right)$ satisfies the minimum extrapolation principle.

Proposition 2 : *$\widehat{I}^m\left(y_t^m\right)$ satisfies Axioms 1-4. Moreover, for any $I^m\left(y_t^m\right)$ that satisfies Axioms 1-4, we have that $\widehat{I}^m\left(y_t^m\right) \subseteq I^m\left(y_t^m\right)$.*

Thus, any set $I^m\left(y_t^m\right)$ that satisfies the Axioms 1-4 also contains $\widehat{I}^m\left(y_t^m\right)$. Putting it differently, by its very construction $\widehat{I}^m\left(y_t^m\right)$ gives an inner bound approximation of the true (but unobserved) input requirement set $I^m\left(y_t^m\right)$ (under the stated technology axioms). This will obtain a specific interpretation for the corresponding technical efficiency measure, as we discuss next.

3.3 Duality between technical and cost efficiency

Given the set $\widehat{I}^m\left(y_t^m\right)$, we can define the following measure of technical input efficiency:

$$\widehat{TE}_t = \min\{\theta |\ \forall m : (\theta \mathbf{q}_t^m, \theta \mathbf{Q}_t) \in \widehat{I}^m\left(y_t^m\right)\}. \tag{9}$$

This measure has a directly similar interpretation as TE_t in (7). The sole difference is that it uses the set $\widehat{I}^m\left(y_t^m\right)$ instead of the true (but unobserved) set $I^m\left(y_t^m\right)$. By construction, we have

$$0 \leq \widehat{TE}_t \leq 1.$$

Next, the result in Proposition 2 directly yields the following relation between $\widehat{TE}_t$ and TE_t:

$$\widehat{I}^m\left(y_t^m\right) \subseteq I^m\left(y_t^m\right) \Rightarrow \widehat{TE}_t \geq TE_t,$$

i.e. the measure $\widehat{TE}_t$ defines an upper bound for TE_t. Putting it differently, $\widehat{TE}_t$ can be interpreted as a "conservative" estimator of the true (but, again, unobserved) technical (in)efficiency of DMU t: it captures equiproportional input reduction that is *certainly* feasible (provided that Axioms 1-4 hold).

Combining (8) and (9), we obtain that the measure $\widehat{TE}_t$ can be computed by solving the next linear programming problem (**LP-3**):

$$\begin{aligned}
&\widehat{TE}_t = \min_{\theta_t \geq 0, \lambda_s^m \geq 0} \theta_t \\
&s.t. \\
&\text{(D-1) } \forall m : \textstyle\sum_{s \in D_t^m} \lambda_s^m \mathbf{Q}_s \leq \theta_t \mathbf{Q}_t \\
&\text{(D-2) } \forall m : \textstyle\sum_{s \in D_t^m} \lambda_s^m \mathbf{q}_s^m \leq \theta_t \mathbf{q}_t^m \\
&\text{(D-3) } \forall m : \textstyle\sum_{s \in D_t^m} \lambda_s^m = 1
\end{aligned}$$

As explained above, θ_t measures the efficiency of DMU t as an equiproportionate reduction of the inputs; the specificity of our efficiency measurement model is that it simultaneously accounts for joint inputs (see constraint (D-1)) and output-specific inputs (see constraint (D-2)). Similar to standard DEA models, the benchmark input vectors are constructed as (convex) combinations of existing DMUs, with every variable λ_s^m representing the weight of each DMU s. In common DEA terminology, the variables λ_s^m are referred to as intensity variables. We note that problem LP-3 obtains separate benchmark (joint and output-specific) input vectors ($\sum_{s \in D_t^m} \lambda_s^m \mathbf{Q}_s$ and $\sum_{s \in D_t^m} \lambda_s^m \mathbf{q}_s^m$) for every different output m. This feature relates to the particular nature of our approach, which explicitly recognizes that each different output is characterized by its own production technology (and, for that reason, its own benchmark input).

Interestingly, the technical efficiency measure $\widehat{TE}_t$ is dually equivalent to the cost efficiency measure $\widehat{CE}_t$ that we introduced in Section 2. In particular, it is easy to verify that problem LP-3 is dual to problem LP-2. Thus, the duality theorem of linear programming directly implies the following result.

Proposition 3 : *We have* $\widehat{TE}_t = \widehat{CE}_t$.

This defines a specific dual interpretation of our multi-output cost efficiency measure in terms of multi-output technical efficiency. As we have explained, this duality result can directly be interpreted in terms of the distinction between multiplier and envelopment DEA models. As such, it clearly suggests our efficiency assessment methodology as extending the existing DEA methodology by explicitly recognizing the multi-output nature of production (in terms of output-specific technologies that are interdependent through jointly used inputs).

As a final remark, we indicate that the result in Proposition 3 may seem paradoxical to some, because our cost and techical efficiency measures are defined under different technology axioms: our definition of the measure $\widehat{CE}_t$ (only) assumes that input requirement sets are nested (Axiom 1), whereas our definition of $\widehat{TE}_t$ additionally uses that these sets are monotone and convex (Axioms 2 and 3). The explanation for this paradox is that monotonicity and convexity of input sets do not interfere with cost efficiency evaluations: for given input prices, minimal cost values remain unaffected after imposing convexity or monotonicity on the input sets.[13] As such, we obtain exactly the same cost efficiency conclusions when imposing these technology properties as when not imposing them. Or, putting it differently, the dual relationship in Proposition 3 would remain intact if we had defined the measure $\widehat{CE}_t$ with respect to input sets that are monotone and convex (in addition to nested), precisely because the values of this cost efficiency measure would not have changed.

3.4 Extensions

Similar to Section 2, we conclude by considering several extensions of our multi-output framework that can build on its technical efficiency (or envelopment) formulation as elaborated above. In particular, drawing on the existing DEA literature, we discuss alternative ways to include additional structure in the efficiency evaluation, so as to obtain a strengthened efficiency evaluation.

First, at the end of Section 2, we indicated the possibility to include restrictions on the shadow prices when computing the cost efficiency measure $\widehat{CE}_t$ on the basis of problem LP-2. Next, using a procedure outlined by Cook, Habadou and Teunter (2000), we showed how one can deal with unobserved output-specific input vectors in practical applications of problem LP-2. Clearly, given that the problems LP-2 and LP-3 are dual to each other, it is also possible to include corresponding restrictions in the calculation of the technical efficiency measure $\widehat{TE}_t$.[14]

Next, our preceding analysis (only) assumed that input requirement sets are nested, monotone and convex. Actually, when adopting an envelopment orientation, DEA applications often use additional production assumptions. Most notably, such additional assumption relate to the nature of the returns-to-scale (i.e. constant, decreasing or increasing) that prevail in the production process.[15] For compactness, we abstract from an exhaustive discussion of such additional assumptions here. However, it is worth emphasizing that these assumptions can be incorporated into our method. More specifically, alternative returns-to-scale assumptions can be implemented in LP-3 by including additional (linear) restrictions defined by Bogetoft (1996). This author focused on (DEA-based) linear programming problems that are formally close to our problem LP-3. Actually, the only difference between our envelopment DEA setting and Bogetoft's original setting pertains exactly to our explicit modeling of specific production technologies for individual outputs.[16] But this difference

[13] See, for example, Varian (1984) for a formal treatment. In particular, this author showed that monotonicity and convexity are nontestable technology properties under cost minimizing production behavior. Essentially, this nontestability result follows from the definition of cost as a linear function of input quantities (evaluated at nonnegative input prices).

[14] See also Podinovski (2004) for a discussion on incorporating weight restrictions in (dual) DEA problems such as LP-3.

[15] See, for example, Cooper, Seiford and Tone (2000) for a detailed discussion of returns-to-scale assumptions that are frequently used in DEA applications.

[16] More specifically, Bogetoft also considers empirical feasibility sets that are constructed as convex-montone hulls based on observed input vectors. In fact, one can verify that in the case of a single output (i.e. $M = 1$) our setting coincides with the one of Bogetoft.

does not interfere with the applicability of Bogetoft's extensions in our multi-output framework. An attractive feature of our multi-output framework is that it allows for invoking different returns-to-scale assumptions for different outputs m.

In Appendix B, we illustrate the possible adaptation of technology assumptions used in more conventional DEA analysis by integrating the assumptions underlying the original DEA model of Charnes, Cooper and Rhodes (CCR, 1978) within our framework. We choose these assumptions because the CCR model is often used as a benchmark model in DEA practice. In addition, the CCR model involves some specific convexity and returns-to-scale assumptions, which allows us to demonstrate the possible use of these assumptions in our framework. Appendix B also compares the resulting input efficiency measure for our multi-output methodology with the measure generated by the basic CCR model. This provides a particular illustration of the fact that explicitly recognizing the multi-output nature of production (by using output-specific technologies) entails an efficiency analysis with greater discriminatory power: when using the same production axioms, our multi-output methodology implies an efficiency measure that generally lies below the standard CCR input efficiency measure, which means that it has greater ability to identify inefficient production behavior.

4 Application

In this section, we demonstrate the practical applicability of our methodology by means of a unique data set collected from the Activity Based Costing (ABC) system of a large service company. As an introduction, we first discuss how ABC systems allow for defining output-specific and joint inputs. Then, we introduce our data and present our efficiency results. We conclude our application by illustrating the managerial implications of our newly developed methodology. Throughout, we will focus on the (primal) cost efficiency representation of our methodology, which we introduced in Section 2. A main motivation is that this representation allows for decomposing (multi-output) cost efficiency in terms of output-specific cost efficiencies (see our discussion of (6)), and this decomposition will be particularly relevant when considering the managerial implications of the efficiency assessment.

At this point, it is worth to recall that in Sections 2.5 and 3.4 we mentioned different possible extensions of our methodology. One such extension will be useful for our following application. Specifically, as we will indicate, our cost efficiency analysis will evaluate the inputs by shadow prices that are subject to price constraints (so excluding unrealistic input prices). The other methodological extensions are not directly relevant for the current application. For example, our use of ABC data makes that we observe the output-specific input vectors, and so we do not need the procedure of Cook, Habadou and Teunter (2000) and Cook and Habadou (2001) (which we discussed in Section 2.5). Next, for the production setting of the service company under evaluation, we are not aware of any a priori argument or empirical evidence that motivates a specific (constant, decreasing or increasing) returns-to-scale assumption. Therefore, we choose not to impose such an assumption here, to avoid the risk of distorting our efficiency analysis by using a wrong (and unverifiable) production assumption. However, it should be clear from our discussion in Section 3.4 that it is actually possible to include specific returns-to-scale assumptions (in applications where such assumptions can be convincingly motivated).

4.1 Input allocation with ABC data

For important strategic decisions such as pricing, product development and product mix decisions, managers need information about the inputs that are consumed by the different outputs. As the allocation of inputs to outputs is not perfectly observable, managers have to rely on the way in which the costing system allocates inputs to outputs. Activity Based Costing (ABC) is a well-known costing methodology and has gained popularity during the last two decades. In an ABC analysis, inputs (or costs) are first allocated to activities by means of resource drivers. In a second stage, the cost of the activities is allocated to the outputs (or products) by means of activity drivers. The selection of resource drivers, activities, and activity drivers is based on a detailed analysis of the production process. Figure 1 presents a graphical representation of an ABC system.

-Insert Figure 1 about here -

In an ABC system, outputs can be considered as consumers of activities and activities can be considered as consumers of inputs. This implies that outputs can be entirely written in terms of activities and in terms of inputs. Thus, by relying on ABC systems, we know how much of an input is used for the production of a certain output. In other words, ABC systems generate information about the input decomposition, which is the distinguishing feature of our newly developed methodology.

Although other costing methodologies also provide information about the allocation of inputs to outputs, there are at least three reasons why our newly developed methodology is more complementary with ABC than with other costing methodologies.[17] First, ABC systems are especially useful for complex production processes with multiple inputs and multiple outputs (see Cooper and Kaplan (1988, 1998)). Such complex production processes are also the focus of our newly developed methodology. Second, it has been shown that, in case of complex production processes, ABC systems are more accurate and lead to better decisions than other costing methodologies (see, for instance, Cardinaels, Roodhooft and Warlop (2004)). Specifically, the inclusion of activities leads to a better approximation of the underlying production process resulting in more accurate information about the allocation of inputs to outputs without having to rely on (unverifiable) assumptions regarding the production technology.[18] Third, ABC systems also enable us to distinguish between output-specific inputs and joint inputs. Specifically, while ABC systems provide a way to allocate inputs to outputs, ABC systems also recognize that some inputs cannot be allocated to the outputs in an accurate way because they lack a direct relationship with the activities

[17] Other costing methodologies are mainly volume-based rather than activity-based. In volume-based costing systems, the inputs (or costs) are allocated to the outputs based on the produced quantities of the different outputs. For instance, if the quantity of an output is 70% of the total quantity of outputs, 70% of the costs are allocated to that output.

[18] Remark that ABC systems do not provide perfect information about the decomposition of inputs to outputs (see, for example, Datar and Gupta (1994) and Labro and Vanhoucke (2007)). Perfect information is not available in general or not available at a reasonable cost. ABC systems are considered as the most accurate approximation of the decomposition of inputs to outputs (see, for example, Bhimani, Horngren, Datar, and Foster (2007)).

and outputs (see Cooper and Kaplan (1991, 1998)) . Such inputs are called "facility-level" inputs. The compensation of the DMU management and the costs for maintenance of the buildings of the DMU are examples of "facility-level" inputs. As efficiency assessments can be biased by allocating inputs that have no cause-and-effect relationship with the outputs, it is appropriate to consider "facility-level" inputs as joint inputs in the efficiency assessment.

4.2 Data

Our empirical application uses data from a large service company active in a European country. It delivers its services to the end customer through 290 offices (i.e. DMUs) that are spread among the country. The offices only differ from each other in terms of their size, which is linked to the size and the population density of the geographical area they operate in. Further, all 290 offices can deliver the same 7 standardized outputs to the end customer, with the corresponding output targets exogenously given (i.e. DMU managers do not have control over the output quantities). As a result, the goal of each office is cost minimization for a given output, which complies with the cost-oriented approach of our methodology.

The company under investigation has its own ABC system, which is implemented at the office level. This implies that we have information about the inputs, resource drivers, costs of activities, activity drivers and outputs for each DMU. The ABC system consists of 7 inputs (i.e. cost categories), 7 activities and 7 outputs. Each DMU uses three types of inputs: labor, transport, and other overhead costs. More specifically, the model contains 3 categories of labor, 3 categories of transport, and 1 category of other overhead costs. The labor and transport subcategories differ from each other in terms of their relationship with the activities. We treat them as distinct inputs because pooling heterogeneous cost categories can decrease the accuracy of the costing system (Labro and Vanhoucke 2007). Labor categories consist of the wages paid to different types of employees. Transport expenditures are fuel costs, maintenance costs and depreciation for different types of vehicles. Other overhead costs consist of all other expenditures made at the DMU level such as pay of the DMU manager, maintenance of the building,... These overhead costs are a typical "facility-level" input (see Section 4.1) and will be treated as a joint input. For each DMU, we obtained expenditure data for every input. Specifically, we treat expenditures as aggregate input quantity indices (i.e. quantities multiplied by prices, with price differences correcting quality differences in the quantity composition). Due to confidentiality and strict Non-Disclosure Agreements, we cannot provide details on the activities, which cover the entire production process of the DMUs, and outputs of the ABC system.

Panel A of Table 1 provides descriptive statistics for the 7 inputs. The large difference between the minimum and maximum values of the different inputs reflects a large variation across the 290 DMUs. We should also mention that some DMUs do not use some of the inputs 4, 5, 6 and 7. Based on the mean relative weights, we can conclude that inputs 1, 2 and 7 are most important. Panel B of Table 1 gives summary statistics for the activities. Activities 3 and 4 are the most input consuming activities. Panel C of Table 1 shows the summary statistics for the outputs. Output 1 is the most important output and takes an average share of 90,78%. The other outputs seem to be far less important. At this point, however, we note that it would be misleading to only consider this output in our efficiency analysis, as it is shrinking in volume year after year and the management is explicitly focusing its attention towards the other outputs.

-Insert Table 1 about here -

Taken together, this empirical application is well suited for demonstrating the practical usefulness of our newly developed efficiency measurement methodology: ABC data are available at the office level, all offices work in a standardized way, which makes them comparable (i.e. DMUs operate in a similar technological environment), offices are quite heterogeneous in terms of inputs used and outputs produced, and cost efficiency (i.e. cost minimization for given output targets) is an appropriate efficiency concept.

4.3 Efficiency results

Our empirical exercise considers four different efficiency measurement models: the first three models involve different specifications of the output-specific and joint inputs for calculating the (multi-output) cost efficiency measure presented in Section 2 ($\widehat{CE}_t$); the fourth model uses a standard cost efficiency measure (SCE_t, which we define below) and will be used as a benchmark model. In each model we use shadow prices to evaluate the different inputs. In doing so, we employ shadow price restrictions to exclude unrealistic input prices; these restrictions have been specified in consultation with the management.

Our first three models solve the problem LP-2 (complemented with linear shadow price restrictions) to compute $\widehat{CE}_t$ for each DMU t under different selections of the output-specific and joint inputs. The first model (*BASIC*) is our core model and considers 6 output-specific inputs (input 1-6) and 1 joint input (input 7). We classified input 7 as a joint input as this input is a facility level input (see Section 4.1). The interpretation of this model is that the ABC system allows us to allocate 6 inputs directly to the outputs, while one input cannot be allocated to any specific output (and, thus, is "shared" by the different outputs). In the second model (*ALL_ALLOCATED*) input 7 is also allocated to the outputs. By contrast, in our third model (*NONE_ALLOCATED*) we do not use any information provided by the ABC system and, thus, all inputs are treated as joint inputs. This model broadly coincides with the model of Cherchye, De Rock and Vermeulen (2008).

We believe that it is useful to compare our findings for these three models with a "standard" cost efficiency measurement model, which does not consider jointly used inputs and/or inputs allocated to specific outputs. This benchmark model (*BENCHMARK*) is defined as follows for each DMU t:[19]

$$SCE_t = \left[\frac{c_t}{\sum_{m=1}^{M} \mathbf{p}_t^{\prime} \mathbf{q}_t^m + \mathbf{P}_t^{\prime} \mathbf{Q}_t} \middle| c_t = \min_{s|\mathbf{y}_s \geq \mathbf{y}_t} \sum_{m=1}^{M} \mathbf{p}_t^{\prime} \mathbf{q}_s^m + \mathbf{P}_t^{\prime} \mathbf{Q}_s \right] \tag{10}$$

[19] For simplicity, we define the standard cost efficiency measure without shadow prices. Including shadow prices proceeds analogously as before. See Cherchye and Vanden Abeele (2005) for a more detailed discussion of this cost efficiency measure. These authors also provide a linear programming formulation to compute the measure when using shadow prices for evaluating the inputs. As indicated above, we will use this shadow price formulation in our empirical exercise, in which we will include the same shadow price restrictions as for the other three efficiency measurement models.

Thus, for given prices $\mathbf{p}_t$ and $\mathbf{P}_t$ this measure divides the minimal cost c_t by the actual cost $\sum_{m=1}^{M}\mathbf{p}_t^{'}\mathbf{q}_t^m + \mathbf{P}_t^{'}\mathbf{Q}_t$; the minimal cost is defined over all DMUs s that produce at least the same amount as DMU t of all different outputs (i.e.$\mathbf{y}_s \geq \mathbf{y}_t$). The essential difference between the measure SCE_t and our multi-output cost efficiency measure $\widehat{CE}_t$ is that this last measure accounts for (interdependent) output-specific production technologies; this complies with the fact that $\widehat{CE}_t$ is composed of output-specific cost efficiency measures $\widehat{CE}_t^m$ (see 6, except that now we decompose $\widehat{CE}_t$ instead of CE_t). In turn, this implies that the newly proposed measure $\widehat{CE}_t$ generally has more discriminatory power than the standard measure SCE_t (because $\widehat{CE}_t$ incorporates more prior information about the underlying production process). We will illustrate this last point in our empirical results.

For each of the four models, we consider two efficiency assessment exercises. In Section 4.3.1, we present the efficiency results without controlling for exogenous variables that may have an impact on DMU efficiency and without correcting for possible outlier behavior of particular DMUs. Subsequently, Section 4.3.2 reports on a second exercise, in which we control for population density as a relevant exogenous variable and simultaneously account for the possibility of outlier behavior. Our selection of population density as the (sole) exogenous variable that is controlled for is the result of consultation with the company management. Next, explicitly accounting for outlier behavior should lead to efficiency results that are more robust (e.g. with respect to measurement errors for inputs and outputs, and non-comparability of DMUs due to (unobserved) heterogeneity of the production environment). In this second exercise, we make use of a probabilistic method that has recently been proposed in a DEA context and that is extensively discussed by Daraio and Simar (2007).[20] This also shows that our new DEA-based methodology can be easily combined with this probabilistic method (as well as with other existing DEA methodologies).

4.3.1 Without control for exogenous variables or outlier behavior

Panel A of Table 2 reports the results for the four efficiency models without control for any exogenous variable and without correction for outlier behavior in the data. Considering the results for the *BASIC*-model, we find that only 10% of the DMUs are efficient and that the average cost reduction potential amounts to 20%. This last result implies that the average office can produce the same output with 20% fewer costs. The results for the *BASIC*-model also show that this model has considerable discriminatory power. This is an interesting property of our methodology, especially when taking into account the attractive structure of the model (with reasonable behavioral assumptions and minimal (unverifiable) production assumptions; see Section 2). In economic terms, our results suggest that, at the aggregate company level, the same output can be produced after reducing totals costs with 123.243.958 EUR. As yet another point of reference, such a cost decrease would imply an increase of the company's EBIT (i.e. earnings before interest and tax) of as much as 33%, ceteris paribus.

The results for the *ALL_ALLOCATED*-model, which are qualitatively similar to the results for the *BASIC*-model, show that 14% of the DMUs are efficient and the average DMU can produce the same output with a cost reduction of 11%. The small differences

[20] The original ideas of this method were presented in Cazals, Florens and Simar (2002) and Daraio and Simar (2005).

between the results for the *ALL_ALLOCATED*- and *BASIC*-models should not be too surprising as the only difference between the two models lies in the treatment of input 7, which accounts for only one sixth of the total costs (i.e. input 7 is considered as a joint input in the *BASIC*-model and an output-specific input in the *ALL_ALLOCATED*-model). However, the differences between both models are substantial enough to make clear that the classification of an input as joint or output-specific matters for the efficiency analysis and for the conclusions that are drawn from it. As for this particular application, we prefer to focus on the *BASIC*-model (and, thus, to treat input 7 as a joint input) as this model better reflects the particular environment of the company.

Next, the empirical results for the *NONE_ALLOCATED*-model are consistent with our expectations. As this model puts very little prior structure by treating all inputs as joint inputs (i.e. no input is specifically allocated to the outputs), we may reasonably expect that the model will have low explanatory power. The results show that almost 90% of the DMUs is declared efficient and the average cost reduction potential is only 2%.

Finally, we consider the results of the *BENCHMARK*-model, which uses the standard cost efficiency measure defined in (10). We find that this model has very low discriminatory power for the given data set: almost all DMUs are efficient. Comparing these findings with our results for the *BASIC*- and *ALL_ALLOCATED*-models provides a strong empirical argument pro using our newly proposed method: the explicit distinction between output-specific and joint inputs in the efficiency assessment does substantially contribute to the discriminatory power of the analysis. In turn, this also pleads for using detailed cost accounting data (generated by an ABC system), which effectively enables such a distinction.

4.3.2 With control for population density and outlier behavior

As mentioned earlier, we also computed efficiency results for the same four models when treating population density as an exogenous variable impacting DMU efficiency, and while accounting for outlier behavior. To this end, we combined our method with the probabilistic order-alpha method of Daouia and Simar (2007). We refer to Daouia and Simar (2007) for a detailed treatment of the method, and restrict to sketching the main idea. The probabilistic method starts by estimating a nonparametric kernel density function through the values of the exogenous variable Z (in our case population density), using a bandwidth h that is determined by cross-validation techniques. Then, it restricts the set of potential comparison partners for each DMU t (with value Z_t for the exogenous variable) to those DMUs of which the corresponding Z value lies within the range $[Z_t - h, Z_t + h]$; as a result, DMU t will only be compared to other DMUs that have a Z value close to Z_t.[21] Specifically, the method repeatedly draws random subsamples (with replacement) from this restricted set of potential comparison partners. For each draw it computes DMU t's cost efficiency, defining a subsample-specific efficiency value. The outlier-robust efficiency measure is then calculated as the average (over all draws) subsample-specific efficiency values. The following efficiency

[21] To be precise, to compute CE_t , the original set of comparison partners for each output m (with corresponding CE_t^m in (6)) is the set of DMUs s such that $y_s^m \geq y_t^m$. Similarly, the original set of comparison partners to compute SCE_t is the sets of DMUs s such that $\mathbf{y}_s \geq \mathbf{y}_t$ (see (10)). The restricted sets of comparison partners contain those DMUs s that additionally satisfy the requirement that Z_s lies within $[Z_t - h, Z_t + h]$.

results pertain to this robust measure (for all four efficiency measurement models under consideration).[22]

Panel B of Table 2 summarizes our findings. A first observation is that the average efficiency value and the number of efficient DMUs for the *BASIC*-model are substantially higher than the corresponding values in Panel A of the same table (i.e. without control for population density and outlier behavior). This suggests that differences in population density as well as outlier behavior may have an important influence on the efficiency results. However, even if we control for these factors, our *BASIC*- model still has a lot of discriminatory power. Specifically, 66,55% of the DMUs are identified as cost inefficient and the mean cost reduction potential still amounts to 6%. The economic impact of this result is still significant: at the aggregate company level, a potential cost reduction of 36.973.187 EUR could be realized without decreasing the output level. Such a cost reduction would increase the EBIT with 10%, ceteris paribus.

Next, we find that the discriminatory power also decreased for the *ALL_ALLOCATED*- and *NONE_ALLOCATED*-models compared to the same models without control for population density and outlier behavior. Generally, the results for these models yield the same qualitative conclusions as before. First, we observe some differences between the results for the *ALL_ALLOCATED*- and *BASIC*- models, which indicates that treating input 7 as a joint input matters for the analysis. Second, we observe that the discriminatory power of the *NONE_ALLOCATED*-model is very low, which again shows that using information about the allocation of the output-specific inputs may substantially enhance the efficiency analysis.

Finally, the *BENCHMARK*-model with a control for population density loses all discriminatory power. If one were to use this method only, it would seem as all offices are operating efficiently. Once again, this result provides a strong empirical argument for using our newly developed efficiency measurement method.

4.4 Managerial Implications

Companies often have multiple business units or offices (i.e. DMUs) that produce identical outputs. A major task of top management is to monitor the efficiency of these DMUs in converting inputs into outputs, and to take appropriate decisions based on the efficiency assessment. Examples of such decisions are evaluation of business unit managers and the linked bonus payments, the installment of benchmarking programs and initiation of improvement actions for bad performing business units, and potentially the dismissal of business unit managers or closing of bad performing business units.

However, accurate efficiency assessment is a complex task for several reasons. First, production processes with multiple outputs are typically characterized by inputs that can be directly allocated to the specific outputs (output-specific inputs) as well as inputs that simultaneously assist in the production of different outputs (joint inputs). The labor cost of employees of a department of a typical supermarket store, for instance, can be directly attributed to the products of that department. The salary of the store manager, however, cannot be attributed to a product or product group. The existence of joint inputs thus

[22] In our exercise, we conducted 200 random draws for calculating these robust measures. In each draw, the number of observations in the subsample equaled 80% of the number of observations in the restricted set of potential comparison partners (where we round to the first higher integer if necessary).

necessitates the use of a method that allocates the joint inputs to the multiple outputs in a way that does not bias the efficiency assessment at the disadvantage of the business unit. Second, business units do not produce the same output mix. Third, even standardized business units operate in different environments. They are subject to different environmental (i.e., exogenous) factors that are beyond their control but influence their efficiency (e.g. population density, average household income,...). As business units should only be held accountable for their inefficiency resulting from controllable factors, and not for the influence of the environment they operate in, a refined methodology is necessary.

We believe that our newly developed methodology has some unique benefits that can improve efficiency assessments of business units and, as a consequence, firm performance. A first benefit is that including information about the allocation of the output-specific inputs to the different outputs substantially improves the discriminatory power of the efficiency assessment. Evidently, an efficiency measurement methodology with more discriminatory power has a greater managerial relevance, as DMU managers can only be motivated to initiate improvement actions if their DMU is identified as inefficient by the efficiency assessment. Furthermore, by treating some inputs as joint inputs and by allocating these inputs to the outputs in a way that does not harm the efficiency result of the particular business unit, our methodology calculates efficiency in a conservative way and takes into account the particular features of the production process. Finally, our methodology can be easily combined with well-known extensions of DEA-based efficiency assessments (e.g. to control for exogenous factors and outlier behavior) so that the benefits of these extensions also pertain to our methodology. Taken together, assessing the efficiency of DMUs by means of our methodology will make the results of the efficiency assessment more acceptable for business unit managers, lead to more improvement actions and, consequently, higher realized cost reductions and improved firm performance.

Another interesting feature of our methodology is that it allows us to decompose the overall efficiency value of a DMU in output-specific efficiency values and corresponding weights (revealing the importance of each individual output in the overall efficiency value; see our discussion of (7) in Section 2). Such a decomposition can lead to more focused improvement actions compared to approaches that do not decompose the overall efficiency value. Indeed, without a decomposition of the overall efficiency value, managers of multi-output DMUs have no clear guidance in terms of the outputs on which they should focus in order to correct the inefficiency that is detected. Taken together, the main distinguishing features of our methodology pertain to the identification of inefficient DMUs and to the fact that it provides managers with more guidance for the installment of improvement actions.

To show the practical usefulness of the decomposition of the overall efficiency value of a DMU, we provide a specific example taken from our application. Panel C of Table 2 reports the output-specific efficiencies and the output weights for three DMUs (A, B and C) that attain the same overall efficiency score (i.e. 0,65). The level of the overall efficiency value indicates that each DMU can produce the same combination of outputs with a cost level that is 35% below the current cost level. While standard methods for efficiency assessment, which typically do not decompose the overall efficiency measure, would stop here, our methodology allows us to go further by analyzing the sources of this cost inefficiencies at the individual output level.

Careful inspection of the output-specific efficiencies reveals some notable differences across the three DMUs. Output 1, for instance, is produced efficiently in DMU B, while DMUs A and C turn out to be inefficient in the production of this output. Considering the weights for output 1 shows that this output is more important for DMU A (i.e. a weight of 0,51) than for DMU B and C (i.e. a weight of respectively 0,11 and 0,06). Summarizing, this example shows that output-specific efficiencies and the corresponding weights can vary a lot between DMUs, which emphasizes the importance of providing this information to managers in order to help them to increase the efficiency of their DMUs.

When considering the other outputs of the DMUs in more detail, we find that the focus of the improvement actions may substantially vary across the three DMUs. For example, DMU B is performing quite well for outputs 4 and 5 (with output 4 much more important than output 5). By contrast, its cost efficiency is much lower for output 3, which is almost as important as output 4. However, the most problematic is output 6, which is only slightly less important than output 3, but has dramatically low efficiency. We also note that the efficiency of outputs 2 and 7 is low, but these outputs are only marginally important for the cost efficiency of DMU B. Taken together, our advice for DMU B is to focus mainly on the production of output 6 and, to a somewhat lesser extent, output 3.

A similar analysis for DMUs A and C yields the following conclusions. First, DMU A can improve its overall efficiency by focusing on output 1, which is very important and is characterized by a potential cost reduction of 14%. In addition, this DMU can fruitfully focus on a more efficient production of outputs 2 and 3, which are a bit less important but characterized by much more room for improvement than output 1. Finally, DMU C should in particular concentrate on output 6, which is both highly important and produced quite inefficiently.

-Insert Table 2 about here -

5 Conclusion

Companies often have multiple business units in which the same outputs are produced. Well-known examples of such companies are Walmart, Home Depot and Mc Donald's. An assessment of the efficiency of the different business units is necessary to manage such companies in an adequate way. This study develops a new DEA-based methodology that improves the efficiency measurement of multi-output DMUs and provides guidance for the improvement actions to restore inefficiency. The distinguishing feature of our methodology is that we include information about the decomposition of the inputs to the outputs. Interestingly, companies often have such information available in their ABC systems.

This new approach to efficiency measurement enriches the production efficiency analysis in two different ways. First, including information about the input decomposition substantially improves the discriminatory power of the efficiency assessment. Specifically, our new methodology is better able to detect productive inefficiencies, which should lead to more improvement actions and higher realized cost savings. A second interesting contribution of our method is that it allows for decomposing the overall efficiency in output-specific efficiencies. Overall cost efficiency measures indicate how well a particular DMU performs in the aggregate, but it does not generate any direct guidance as to which actions can effectively improve the observed inefficiencies. By contrast, output-specific efficiency measures effectively identify the outputs on which DMUs should focus to remedy the observed inefficiency. Given that business units typically have limited resources to remedy inefficiencies, our methodology helps to better allocate these scarce resources to the outputs that contribute the most to the inefficiency that is observed. Summarizing, our methodology will lead to more improvement actions as well as more focused improvement actions.

We see multiple avenues for follow-up research. First, at a methodological level, our approach allows for a richer type of efficiency analysis, because it explicitly recognizes that different outputs are characterized by own production technologies that may be interdependent because of jointly used inputs. As we indicated at the end of Section 2.1, evaluating the joint inputs by using output-specific implicit prices (that add up to the observed prices) corresponds to a Pareto efficient use of these inputs. Instead of Pareto efficiency, one may also assume a Nash equilibrium allocation for multi-output production (which need not necessarily be Pareto efficient). Here, one may fruitfully build on Cherchye, Demuynck and De Rock (2011), who considered this Nash equilibrium criterion in a formally close consumption setting.[23] More generally, we believe that our modeling of output-specific production technologies opens the way for a whole new spectrum of applications of multi-output efficiency analysis.

Next, as for empirical applications, we have suggested using ABC data to obtain information about the decomposition of the output-specific inputs to the different outputs. In this respect, previous research has shown that the accuracy of ABC systems depends on the characteristics of the economic environment, such as diversity in the resource consumption patterns (Labro and Vanhoucke 2007). Future research could investigate how the accuracy of costing systems and the determinants thereof influence the accuracy of the efficiency assessments. More broadly, we have only scratched the surface in exploring the interface betwen operations research and cost accounting, and the opportunities for future research are enormous.

References

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

[2] Allen, R., A.D. Athanassopoulos, R.G. Dyson, E. Thanassoulis. 1997. Weight Restrictions and Value Judgements in DEA: Evolution, Development and Future Directions. *Annals of Operations Research* **73** 13-34.

[3] Beasley, J.E. 1995. Determining teaching and research efficiencies, *Journal of the Operational Research Society* **46**, 441–452.

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

[23] See also Cherchye, Demuynck, De Rock and De Witte (2011) for related discussion.

[5] Bhimani, A., C.T. Horngren, S. Datar, G. Foster. 2007. *Management and cost accounting*. Financial Times Press

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

[7] Cardinaels, E., F. Roodhooft, L. Warlop. 2004. Customer profitability analysis reports for resource allocation: the role of complex marketing environments. *Abacus* **40** 238-258.

[8] Cazals, S., J.P. Florens, L. Simar. 2002. Nonparametric Frontier Estimation: A Robust Approach. *Journal of Econometrics* **106** 1-25.

[9] Charnes, A., W.W. Cooper. 1962. Programming with linear fractional functionals. *Naval Research Logistics Quarterly* **9** 181–186.

[10] Charnes, A., W.W. Cooper, E. Rhodes. 1978. Measuring the Efficiency of Decision Making Units. *European Journal of Operational Research* **2** 429-444.

[11] Cherchye, L., P. Vanden Abeele. 2005. On Research Efficiency: A Micro-Analaysis of Dutch University Research in Economics and Business Management. *Research Policy* **34** 495-516.

[12] Cherchye, L., T. Demuynck, B. De Rock. 2011. Revealed preference analysis of noncooperative household consumption. *Economic Journal* **121** 1073-1096.

[13] Cherchye, Demuynck, De Rock and De Witte. 2011. Nonparametric analysis of multi-output production with joint inputs. *Working paper.* (Under revision for *Economic Journal.*)

[14] Cherchye, L., B. De Rock, F. Vermeulen. 2007. The collective model of household consumption: a nonparametric characterization. *Econometrica* **75** 553-574.

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

[16] Cherchye, L., B. De Rock, F. Vermeulen. 2011. The revealed preference approach to collective consumption behavior: testing and sharing rule recovery. *Review of Economic Studies* **78** 176-198.

[17] Cherchye, L., W. Moesen, N. Rogge, T. Van Puyenbroeck. 2007. An introduction to 'benefit of the doubt' composite indicators. *Social Indicators Research* **82** 111-145.

[18] Chiappori, P.-A. 1988. Rational household labor supply. *Econometrica* **56** 63-89.

[19] Cook, W.D., M. Habadou, H.J.H Teunter. 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.

[20] Cook, W.D., M. Hababou. 2001. Sales performance measurement in bank branches, *OMEGA* **29**, 299–307.

[21] Cook, W.D., R.H. Green. 2004. Multicomponent efficiency measurement and core business identification in multiplant firms: A DEA model, *European Journal of Operational Research* **157**, 540–551.

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

[23] Cooper, R., R.S. Kaplan. 1988. Measure Costs Right: Make The Right Decisions. *Harvard Business Review* **67** (September-October) 96-105.

[24] Cooper, R., R.S. Kaplan. 1991. Profit priorities from Activity Based Costing. *Harvard Business Review* **70** (May-June): 130-135.

[25] Cooper, R., R.S. Kaplan. 1998. *Cost & Effect: Using Integrated Cost Systems to Drive Profitability and Performance*. Harvard Business School Press.

[26] Cooper, W.W., L.M. Seiford, K. Tone. 2000. *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*. Kluwer Academic Publishers.

[27] Datar, S., M. Gupta. 1994. Aggregation, specification, and measurement errors in product costing. *The Accounting Review* **69** 567-592.

[28] Daouia, A., L. Simar, 2007. Nonparametric efficiency analysis: A multivariate conditional quantile approach. *Journal of Econometrics* **140** 375-400.

[29] Daraio, C., L. Simar. 2005. Introducing environmental variables in nonparametric frontier models: a probabilistic approach. *Journal of Productivity Analysis* **24** 93-121.

[30] Daraio, C., L. Simar. 2007. *Advanced Robust and Nonparametric Methods in Efficiency Analysis. Methodology and Applications*. Springer.

[31] Diewert, W.E., C. Parkan. 1983. Linear Programming Tests of Regularity Conditions for Production Frontiers. In *Quantitative Studies on Production and Prices* (W. Eichhorn, R. Henn, K. Neumann and R.W. Shephard, Eds.), Physica-Verlag.

[32] Färe, R., S. Grosskopf, C.A.K. Lovell. 1994. *Production Frontiers*. Cambridge University Press.

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

[34] Gosselin, M. 2007. A Review of Activity-Based Costing: Technique, Implementation and Consequences. In *Handbook of Management Accounting Research*, edited by C.S. Chapman, A.G. Hopwood, and M.D. Shields. Elsevier.

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

[36] Kuosmanen, T., L. Cherchye, T. Sipiläinen. 2006. The Law of one Price in Data Envelopment Analysis: Restricting Weight Flexibility across Firms. *European Journal of Operational Research* **170** 735-757.

[37] Labro, E., and M. Vanhoucke. 2007. A Simulation Analysis of Interactions among Errors in Costing Systems. *The Accounting Review* **82** 939-962.

[38] Myles, G.D. 1995. *Public Economics*. Cambridge University Press.

[39] Pedraja-Chaparro, F., J. Salinas-Jimenez, P. Smith. 1997. On the Role of Weight Restrictions in Data Envelopment Analysis. *Journal of Productivity Analysis* **8** 215-230.

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

[41] Podinovski V. 2004. Production trade-offs and weight restrictions in data envelopment analysis. *Journal of the Operational Research Society* **55** 1311–1322.

[42] 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.

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

Appendix A: Proofs

Proof of Proposition 1

As a first step, we show that, if DMU t is multi-output cost efficient, then there must exist implicit prices $\mathbf{P}_t^m \in \mathbb{R}_+^{Njoin}$ $(1 \leq m \leq M)$ such that, for each output m, it holds that: if, for some DMU s, $y_s^m \geq y_t^m$, then $\mathbf{p}_t'\mathbf{q}_t^m + (\mathbf{P}_t^m)'\mathbf{Q}_t \leq \mathbf{p}_t'\mathbf{q}_s^m + (\mathbf{P}_t^m)'\mathbf{Q}_s$. Because DMU t is multi-output cost efficient, conditions (CE-1) and (CE-2) in Definition 2 are satisfied. Condition (CE-1) implies, for nested input sets,

$$y_t^m \leq y_s^m \Rightarrow (\mathbf{q}_s^m, \mathbf{Q}_s) \in I^m(y_t^m). \tag{11}$$

Next, (CE-2) is satisfied only if

$$\mathbf{p}_t'\mathbf{q}_t^m + (\mathbf{P}_t^m)'\mathbf{Q}_t = \min_{(\mathbf{q}^m, \mathbf{Q}) \in I^m(y_t^m)} \mathbf{p}_t'\mathbf{q}^m + (\mathbf{P}_t^m)'\mathbf{Q}. \tag{12}$$

Combining (11) and (12), we get

$$y_t^m \leq y_s^m \Rightarrow \mathbf{p}_t'\mathbf{q}_t^m + (\mathbf{P}_t^m)'\mathbf{Q}_t \leq \mathbf{p}_t'\mathbf{q}_s^m + (\mathbf{P}_t^m)'\mathbf{Q}_s,$$

which gives the wanted result.

As a second step, we show that DMU t is multi-output cost efficient (i.e. conditions (CE-1) and (CE-2) in Definition 2 are satisfied) if there exist implicit prices $\mathbf{P}_t^m \in \mathbb{R}_+^{Njoin}$ $(1 \leq m \leq M)$ such that, for each output m, it holds that: if, for some DMU s, $y_s^m \geq y_t^m$, then $\mathbf{p}_t^{'}\mathbf{q}_t^m + (\mathbf{P}_t^m)^{'}\mathbf{Q}_t \leq \mathbf{p}_t^{'}\mathbf{q}_s^m + (\mathbf{P}_t^m)^{'}\mathbf{Q}_s$. To show the result, for each output m and DMU t we construct the input requirement set

$$I^m(y_t^m) = \{(\mathbf{q}^m, \mathbf{Q}) \,|\exists s \in D_t^m : \mathbf{q}_s^m \leq \mathbf{q}^m, \mathbf{Q}_s \leq \mathbf{Q}\}, \qquad (13)$$
$$\text{for } D_t^m = \{s|\ y_t^m \leq y_s^m\}.$$

In words, this input requirement set is constructed as the (positive) monotone hull of the input vectors ($\mathbf{q}_s^m$, $\mathbf{Q}_s$) associated with all DMUs s that produce at least the output y_t^m (i.e. $s \in D_t^m$).[24] By their very construction, these input requirement sets meet condition (CE-1) in Definition 2 (i.e. they are nested and satisfy $(\mathbf{q}_t^m, \mathbf{Q}_t) \in I^m(y_t^m)$).

Next, to show that condition (CE-2) in Definition 2 is satisfied, we note that

$$\min_{s \in D_t^m} \mathbf{p}_t^{'}\mathbf{q}_s^m + (\mathbf{P}_t^m)^{'}\mathbf{Q}_s = \min_{(\mathbf{q}^m, \mathbf{Q}) \in I^m(y_t^m)} \mathbf{p}_t^{'}\mathbf{q}^m + (\mathbf{P}_t^m)^{'}\mathbf{Q}; \qquad (14)$$

this directly follows from the monotone hull construction of $I^m(y_t^m)$ in (13).[25] Because $\mathbf{p}_t^{'}\mathbf{q}_t^m + (\mathbf{P}_t^m)^{'}\mathbf{Q}_t \leq \mathbf{p}_t^{'}\mathbf{q}_s^m + (\mathbf{P}_t^m)^{'}\mathbf{Q}_s$ for any $s \in D_t^m$, and using that $t \in D_t^m$ by construction, the equality in (14) effectively obtains the wanted result, i.e.

$$\mathbf{p}_t^{'}\mathbf{q}_t^m + (\mathbf{P}_t^m)^{'}\mathbf{Q}_t = \min_{s \in D_t^m} \mathbf{p}_t^{'}\mathbf{q}^m + (\mathbf{P}_t^m)^{'}\mathbf{Q} = \min_{(\mathbf{q}^m, \mathbf{Q}) \in I^m(y_t^m)} \mathbf{p}_t^{'}\mathbf{q}^m + (\mathbf{P}_t^m)^{'}\mathbf{Q}.$$

Proof of Proposition 2

We first need to verify that $\widehat{I}^m(y_t^m)$ satisfies Axioms 1-4. For Axioms 2, 3 and 4 the result follows directly from the definition of $\widehat{I}^m(y_t^m)$ (in (8)) as the convex-monotone hull of the DMUs s in the set D_t^m (where we obtain Axiom 4 by noting that, by construction, $t \in D_t^m$). Next, to show that Axiom 1 is satisfied, we note that $y_t^m \geq \overline{y}_t^m$ implies $D_t^m \subseteq \overline{D}_t^m$ (with $\overline{D}_t^m = \{s|\ \overline{y}_t^m \leq y_s^m\}$), which in turn implies $I^m(y^m) \subseteq I^m(\overline{y}^m)$.

It remains to prove that, for any $I^m(y_t^m)$ that satisfies Axioms 1-4, we have $\widehat{I}^m(y_t^m) \subseteq I^m(y_t^m)$. Consider any $(\mathbf{q}_*^m, \mathbf{Q}_*) \in \widehat{I}^m(y_t^m)$. We must prove that $(\mathbf{q}_*^m, \mathbf{Q}_*) \in I^m(y_t^m)$. As a

[24] Here, it is worth to indicate that, in principle, we could also have obtained our result by using the convex-monotone hull construction of the same input vectors ($\mathbf{q}_s^m$, $\mathbf{Q}_s$) (which we consider in Section 3 of the main text; see (8)). However, for the purpose of proving Proposition 2, it suffices to use the monotone hull construction in (13).

[25] In particular, we can use that, for any $s \in D_t^m$, $\mathbf{p}_t^{'}\mathbf{q}_s^m + (\mathbf{P}_t^m)^{'}\mathbf{Q}_s \leq \mathbf{p}_t^{'}\mathbf{q}^m + (\mathbf{P}_t^m)^{'}\mathbf{Q}$ if $\mathbf{q}_s^m \leq \mathbf{q}^m$ and $\mathbf{Q}_s \leq \mathbf{Q}$.

preliminary step, we indicate that, by the definition of $\widehat{I}^m(y_t^m)$,

$$\sum_{s \in D_t^m} \lambda_s^m \mathbf{q}_s^m \leq \mathbf{q}_*^m \text{ and } \sum_{s \in D_t^m} \lambda_s^m \mathbf{Q}_s \leq \mathbf{Q}_*,$$
$$\text{for some } \lambda_s^m \geq 0 \text{ such that } \sum_{s \in D_t^m} \lambda_s^m = 1.$$

We can now prove $(\mathbf{q}_*^m, \mathbf{Q}_*) \in I^m(y_t^m)$ by using that $I^m(y_t^m)$ satisfies the Axioms 1-4. As a first step, we note that Axiom 4 implies

$$(\mathbf{q}_t^m, \mathbf{Q}_t) \in I^m(y_t^m).$$

In combination with Axiom 1, this entails

$$\forall s : y_t^m \leq y_s^m \Rightarrow (\mathbf{q}_s^m, \mathbf{Q}_s) \in I^m(y_t^m),$$

and thus, because $D_t^m = \{s |\ y_t^m \leq y_s^m\}$,

$$\{(\mathbf{q}_s^m, \mathbf{Q}_s) \,| s \in D_t^m\} \subseteq I^m(y_t^m).$$

Combining this with Axioms 2 and 3, we get that

$$(\mathbf{q}^m, \mathbf{Q}) \in I^m(y_t^m) \text{ if } \sum_{s \in D_t^m} \lambda_s^m \mathbf{q}_s^m \leq \mathbf{q}^m \text{ and } \sum_{s \in D_t^m} \lambda_s^m \mathbf{Q}_s \leq \mathbf{Q}$$
$$\text{for some } \lambda_s^m \geq 0 \text{ such that } \sum_{s \in D_t^m} \lambda_s^m = 1,$$

which effectively obtains

$$(\mathbf{q}_*^m, \mathbf{Q}_*) \in I^m(y_t^m).$$

Proof of Proposition 3

The result follows from the fact that problem LP-3 is dual to problem LP-2, using the duality theorem of linear programming.

Appendix B: Relationship with the CCR model

In this appendix, we discuss the relationship between our model for multi-output efficiency analysis and the original DEA model of Charnes, Cooper and Rhodes (CCR, 1978). To compactify our exposition, we will only consider the "envelopment" formulation (as in Section 3) of the CCR model, which is mostly considered in applied DEA work. However, it is worth indicating that our following argument could also be stated in "multiplier" form (as in Section 2).

We will first introduce the technology axioms underlying the CCR model, to subsequently define the corresponding input efficiency measure. In a following step, we then demonstrate how to adapt the same axioms to our multi-output methodology. Here, we will also show that the correspondingly defined input efficiency measure is generally situated below the standard CCR input efficiency measure. This provides a specific illustration of the more general point that an explicit recognition of the multi-output nature of production (by using output-specific technologies) entails an efficiency analysis with greater discriminatory power: when using the same production axioms, the resulting efficiency measures have greater ability to identify inefficient production behavior.

Technology axioms of the CCR model

As explained in the main text, standard DEA models do not consider output-specific input requirement sets, and they do not distinguish between output-specific inputs and joint inputs. Formally, for each DMU t with output vector $\mathbf{y}_t$, DEA models consider an "aggregate" N-vector of inputs $\mathbf{q}_t \in \mathbb{R}_+^N$, which gives the data set

$$S = \{(\mathbf{y}_t, \mathbf{q}_t) \,|t = 1, ..., T\} .$$

Correspondingly, the production technology is represented by input requirement sets $I(\mathbf{y})$, which contain all the input combinations $\mathbf{q}$ that can produce the output vector $\mathbf{y}$:

$$I(\mathbf{y}) = \{\mathbf{q} \in \mathbb{R}_+^N | \mathbf{q} \text{ can produce } \mathbf{y}\}.$$

At this point, it is useful to clarify the relation with the set-up that we used in Sections 2 and 3. In these sections, we divided the N-dimensional input vector $\mathbf{q}$ into M output-specific input vectors $\mathbf{q}^m \in \mathbb{R}_+^{Nspec}$ and a joint input vector $\mathbf{Q} \in \mathbb{R}_+^{Njoint}$. As such, we have $\mathbf{q} = \sum_{m=1}^{M} \mathbf{q}^m + \mathbf{Q}$ and $N = Nspec + Njoint$.

For input requirement sets $I\left(\mathbf{y}\right)$, the CCR model uses the following technology axioms:

Axiom 5 : $\mathbf{y} \geq \mathbf{y}^\star \Rightarrow I(\mathbf{y}) \subseteq I(\mathbf{y}^\star)$.

Axiom 6 : $\mathbf{q} \in I\left(\mathbf{y}\right)$ *and* $\mathbf{q}^\star \geq \mathbf{q} \Rightarrow \mathbf{q}^\star \in I(\mathbf{y})$.

Axiom 7 : $\mathbf{q} \in I\left(\mathbf{y}\right)$ *and* $\mathbf{q}^\star \in I\left(\mathbf{y}^\star\right) \Rightarrow \forall \lambda \in [0,1] : \lambda \mathbf{q} + (1-\lambda)\,\mathbf{q}^\star \in I^m(\lambda \mathbf{y} + (1-\lambda)\,\mathbf{y}^\star)$.

Axiom 8 : $\mathbf{q} \in I\left(\mathbf{y}\right) \Rightarrow \forall \lambda \geq 0 : \lambda \mathbf{q} \in I(\lambda \mathbf{y})$.

Axiom 9 : $(\mathbf{y}_t, \mathbf{q}_t) \in S \Rightarrow \mathbf{q}_t \in I(\mathbf{y}_t)$.

The Axioms 5, 6 and 9 have a directly similar interpretation as the Axioms 1, 2 and 4. The mere difference is that we now consider input requirement sets $I(\mathbf{y})$ defined for the full output vector $\mathbf{y}$, whereas before we regarded input requirement sets $I^m(y_t^m)$ defined for individual outputs m.

Next, the Axioms 7 and 8 are specific to the CCR model and do not have direct counterparts in the main text. First, Axiom 7 imposes convexity in input-output space. In words, if two input-output combinations ($\mathbf{q}$, $\mathbf{y}$) and ($\mathbf{q}^\star$, $\mathbf{y}^\star$) are technically feasible, then every convex combination $(\lambda \mathbf{q} + (1-\lambda)\,\mathbf{q}^\star,\ \lambda \mathbf{y} + (1-\lambda)\,\mathbf{y}^\star)$, with $\lambda \in [0,1]$, is also technically feasible. Second, a specific feature of the CCR model is that it imposes constant returns-to-scale (which is captured by Axiom 8): if the input-output combination $(\mathbf{q}, \mathbf{y})$ is technically feasible, then any positive rescaling $(\lambda\mathbf{q}, \lambda\mathbf{y})$, with $\lambda \geq 0$, is also technically feasible.

It can be shown that, under the minimum extrapolation principle, the Axioms 5-9 define the following empirical estimate of each set $I(\mathbf{y}_t)$:

$$\widehat{I}^{CCR}(\mathbf{y}_t) = \left\{ \mathbf{q} \;\middle|\; \begin{array}{c} \sum_{s\in S} \lambda_s \mathbf{q}_s \leq \mathbf{q}, \\ \sum_{s\in S} \lambda_s \mathbf{y}_s \geq \mathbf{y}_t, \lambda_s \geq 0 \end{array} \right\}.$$

This production set approximation defines the CCR measure of technical input efficiency (**LP-4**):

$$\begin{array}{l} \widehat{TE}_t^{CCR} = \min\limits_{\theta_t \geq 0, \lambda_s \geq 0} \theta_t \\ s.t. \\ \text{(CCR-1)} \sum_{s\in S} \lambda_s \mathbf{q}_s \leq \theta_t \mathbf{q}_t \\ \text{(CCR-2)} \sum_{s\in S} \lambda_s \mathbf{y}_s \geq \mathbf{y}_t \end{array}$$

CCR axioms and output-specific technologies

We can now apply the technology axioms underlying the CCR model to our methodology with output-specific technologies. This requires the following modifications of the Axioms 7 and 8:

Axiom 10 : $(\mathbf{q}^m, \mathbf{Q}) \in I^m(y^m)$ *and* $(\mathbf{q}^{m\star}, \mathbf{Q}^\star) \in I^m\left(y^{m^\star}\right) \Rightarrow \forall\lambda \in [0,1] : \lambda(\mathbf{q}^{m\star}, \mathbf{Q}^\star) + (1-\lambda)(\mathbf{q}^{m\star}, \mathbf{Q}^\star) \in I^m(\lambda y^m + (1-\lambda) y^{m\star})$.

Axiom 11 : $(\mathbf{q}^m, \mathbf{Q}) \in I^m(y^m) \Rightarrow \forall\lambda \geq 0 : \lambda(\mathbf{q}^m, \mathbf{Q}) \in I^m(\lambda y^m)$.

The interpretation of these axioms is readily analogous to before. The only difference is that the Axioms 10 and 11 pertain to input requirement sets defined for individual outputs (i.e. $I^m(y^m)$ for output y^m), whereas the Axioms 7 and 8 relate to input sets defined for producible output vectors (i.e. $I(\mathbf{y})$ for output vector $\mathbf{y}$). Thus, Axiom 10 implies convexity in input-output space for every individual output m. Generally, this axiom is stronger than Axiom 3, which only imposed convexity in input space. Finally, Axiom 11 states that the production of each output m is characterized by constant returns-to-scale.

We can show that, under the minimum extrapolation principle, the Axioms 1, 2, 4, 10 and 11 define the next empirical estimate for the output-specific sets $I(y_t^m)$:[26]

$$\widehat{I}^{m,CCR}(y_t^m) = \left\{ \mathbf{q} \;\middle|\; \begin{array}{c} \sum_{s\in S} \lambda_s^m \mathbf{q}_s^m \leq \mathbf{q}^m, \sum_{s\in S} \lambda_s^m \mathbf{Q}_s \leq \mathbf{Q} \\ \sum_{s\in S} \lambda_s^m y_s^m \geq y_t^m, \lambda_s^m \geq 0 \end{array} \right\}.$$

[26] For compactness, we do not explicitly prove that the Axioms 1, 2, 4, 10 and 11 define the empirical estimate $\widehat{I}^{m,CCR}(y_t^m)$ for $I(y_t^m)$ under the minimum extrapolation principle. But the argument is similar to the one leading up to Proposition 2.

In turn, this defines the "extended" (E) CCR measure for input efficiency (**LP-5**)

$$\begin{array}{l} \widehat{TE}_t^{CCR-E} = \min_{\theta_t \geq 0, \lambda_s^m \geq 0} \theta_t \\ s.t. \\ \text{(CCR-E-1) } \forall m : \sum_{s \in S} \lambda_s^m \mathbf{Q}_s \leq \theta_t \mathbf{Q}_t \\ \text{(CCR-E-2) } \forall m : \sum_{s \in S} \lambda_s^m \mathbf{q}_s^m \leq \theta_t \mathbf{q}_t^m \\ \text{(CCR-E-3) } \forall m : \sum_{s \in S} \lambda_s^m y_s^m \geq y_t^m \end{array}$$

Comparison of LP-4 and LP-5 yields that an explicit recognition of the multi-output nature of production effectively implies a strengthened efficiency evaluation. In particular, we get generally lower input efficiency values when using exactly the same production axioms for the output-specific sets $I^m(y^m)$ (leading to LP-5) as the CCR model uses for the sets $I(\mathbf{y})$ (leading to LP-4). This feature relates to the particular nature of our approach, with each different output characterized by its own production technology. As an implication, LP-5 can identify separate benchmark (joint and output-specific) vectors ($\sum_{s \in S} \lambda_s^m \mathbf{Q}_s$ and $\sum_{s \in S} \lambda_s^m \mathbf{q}_s^m$) for the inputs related to every different output m, while LP-4 is bound to define a single input benchmark ($\sum_{s \in S} \lambda_s \mathbf{q}_s$). It is this greater flexibility to select (output-specific) benchmarks that gives our multi-output methodology more discriminatory power. The following result provides a formal statement.

Proposition 4 : *We have* $\widehat{TE}_t^{CCR-E} \leq \widehat{TE}_t^{CCR}$.

Proof. The result follows directly if we can show that any feasible value of θ_t in LP-4 (supported by a specification of λ_s, $s \in S$) can be reconstructed as a feasible value of θ_t in LP-5 (supported by some λ_s^m, $s \in S$ and $m = 1, ..., M$).

Let θ_t^* and λ_s^* ($s \in S$) be a feasible solution of LP-4. Then, we must show that there exist λ_s^{m*} ($s \in S$ and $m = 1, ..., M$) such that θ_t^* and λ_s^{m*} is a feasible solution of LP-5. We obtain this conclusion for $\lambda_s^{m*} = \lambda_s^*$ (for all m). First, $\lambda_s^{m*} = \lambda_s^*$ obtains that CCR-E-3 reduces to CCR–2. Similarly, the constraint CCR-E-1 reduces to (for any m)

$$\sum_{s \in S} \lambda_s^* \mathbf{Q}_s \leq \theta_t \mathbf{Q}_t.$$

When we add this last constraint to the CCR-E-2 constraints associated with the different outputs m, we get (again using $\lambda_s^{m*} = \lambda_s^*$ in the CCR-E-2 constraints)

$$\sum_{s \in S} \lambda_s^* \left(\sum_{m=1}^{M} \mathbf{q}_s^m + \mathbf{Q}_s \right) \leq \theta_t \left(\sum_{m=1}^{M} \mathbf{q}_t^m + \mathbf{Q}_t \right),$$

or

$$\sum_{s \in S} \lambda_s^* \mathbf{q}_s \leq \theta_t \mathbf{q}_t,$$

where we use that $\mathbf{q} = \sum_{m=1}^{M} \mathbf{q}^m + \mathbf{Q}$. Thus, because θ_t^* and λ_s^* is a feasible solution of LP-4, we conclude that θ_t^* and λ_s^{m*} ($= \lambda_s^*$) is a feasible solution of LP-5. ■

FIGURE 1: ABC model

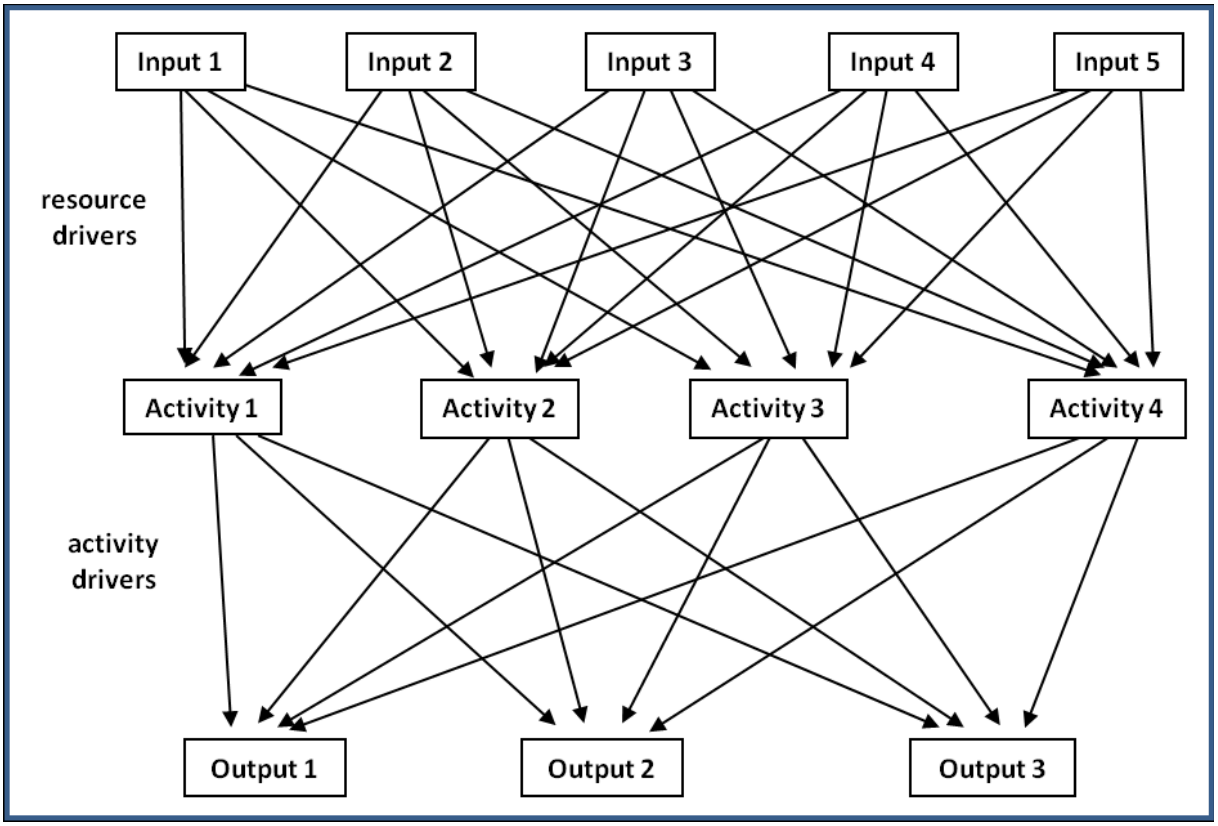

TABLE 1: Summary statistics for input, activities, and outputs

PANEL A: SUMMARY STATISTICS FOR INPUTS

	Input 1	Input 2	Input 3	Input 4	Input 5	Input 6	Input 7
Minimum	2.300,53	6.522,28	2.402,56	0,00	0,00	0,00	0,00
1st Quartile	416.512,75	178.020,86	27.060,68	16.722,70	861,31	99,13	8.224,88
Median	681.166,76	346.869,95	50.897,87	38.132,60	3.913,21	965,87	18.776,13
3rd Quartile	1.235.242,32	641.469,51	96.741,76	69.255,21	10.609,42	5.327,23	84.246,26
Maximum	7.852.652,61	3.106.230,5	807.476,34	512.198,29	50.478,14	555.005,36	5.836.885,8
Mean	1.083.253,08	511.329,46	88.798,65	57.454,91	7.351,60	29.023,07	347.685,06
Stddev	1.256.051,39	493.970,15	113.984,00	68.758,95	9.163,86	76.749,02	873.538,76
Mean Relative Weight	0,51	0,24	0,04	0,03	0,00	0,01	0,16

PANEL B: SUMMARY STATISTICS FOR ACTIVITIES

	Activity 1	Activity 2	Activity 3	Activity 4	Activity 5	Activity 6	Activity 7
Minimum	50.715,01	16.633,58	109.033,49	83.422,97	57.595,20	660,08	28.828,71
1st Quartile	134.417,95	44.086,60	272.126,91	221.109,00	152.653,60	1.749,52	76.409,25
Median	208.236,47	68.297,70	432.590,15	342.535,79	236.486,62	2.710,31	118.371,04
3rd Quartile	365.520,84	119.884,06	763.735,28	601.258,59	415.108,78	4.757,45	207.778,59
Maximum	1.646.721,96	540.094,28	3.239.597,01	2.708.753,15	1.870.122,49	21.432,96	936.071,30
Mean	314.939,21	103.294,23	647.817,50	518.055,03	357.665,06	4.099,10	179.025,71
Stddev	298.032,42	97.749,11	601.898,45	490.244,43	338.464,63	3.879,05	169.415,12
Mean Relative Weight	0,15	0,05	0,30	0,24	0,17	0,00	0,08

PANEL C: SUMMARY STATISTICS FOR OUTPUTS

	Output 1	Output 2	Output 3	Output 4	Output 5	Output 6	Output 7
Minimum	5.647,86	0,00	0,00	11,97	2,06	33,25	2,68
1st Quartile	17.351,95	0,00	65,75	67,88	7,20	117,18	140,65
Median	28.991,22	0,00	1.501,44	99,32	13,69	206,15	240,37
3rd Quartile	48.648,71	0,00	3.240,62	159,19	23,09	369,07	458,22
Maximum	167.844,62	70,00	28.251,79	837,27	78,42	5.515,25	4.171,21
Mean	37.935,31	0,93	2.859,37	134,84	18,22	382,28	456,77
Stddev	30.129,41	5,28	4.540,18	108,95	14,95	574,87	644,00
Mean Relative Weight	0,91	0,00	0,07	0,00	0,00	0,01	0,01

TABLE 2: Efficiency results

PANEL A: EFFICIENCY RESULTS WITHOUT CONTROL FOR POPULATION DENSITY AND OUTLIER BEHAVIOR

Efficiency measure	BASIC	ALL_ALLOCATED	NONE_ALLOCATED	BENCHMARK
Minimum	0,23	0,30	0,52	0,61
1st Quartile	0,68	0,86	1,00	1,00
Median	0,83	0,92	1,00	1,00
3rd Quartile	0,96	0,97	1,00	1,00
Maximum	1,00	1,00	1,00	1,00
Mean	0,80	0,89	0,98	1,00
Stdev	0,18	0,12	0,07	0,03
Efficient DMUs				
Number	29	40	259	285
Percentage	10,00%	13,79%	89,31%	98,28%

PANEL B: EFFICIENCY RESULTS WITH CONTROL FOR POPULATION DENSITY AND OUTLIER BEHAVIOR

Efficiency measure	BASIC	ALL_ALLOCATED	NONE_ALLOCATED	BENCHMARK	POP.DENSITY
Minimum	0,43	0,42	0,59	1,00	5,2
1st Quartile	0,92	0,88	1,00	1,00	16,5
Median	0,98	0,93	1,00	1,00	22,2
3rd Quartile	1,00	0,97	1,00	1,00	34,2
Maximum	1,00	1,00	1,00	1,00	253,5
Mean	0,94	0,91	0,99	1,00	33,9
Stdev	0,10	0,10	0,05	0	38,9
Efficient DMUs					
Number	97	47	270	290	-
Percentage	33,45%	16,21%	93,10%	100%	-

PANEL C: DECOMPOSITION OF OVERALL EFFICIENCY FOR THREE DMUS (weights between brackets)

	Overall	Output1	Output2	Output3	Output4	Output5	Output6	Output7
A	0,65	0,86	0,25	0,56	0,87	0,22	0,00	0,10
		(0,51)	(0,16)	(0,17)	(0,07)	(0,06)	(0,02)	(0,01)
B	0,65	1,00	0,21	0,69	0,94	0,99	0,01	0,01
		(0,11)	(0,02)	(0,23)	(0,29)	(0,10)	(0,21)	(0,04)
C	0,65	0,89	0,40	0,50	0,53	0,84	0,67	1
		(0,06)	(0,05)	(0,09)	(0,09)	(0,04)	(0,67)	(0,00)