Ann Oper Res (2006) 145:51–68
DOI 10.1007/s10479-006-0025-8

An efficiency measurement framework for multi-stage production systems

Boaz Golany · Steven T. Hackman · Ury Passy

Published online: 24 June 2006

Abstract We develop an efficiency measurement framework for systems composed of two subsystems arranged in series that simultaneously computes the efficiency of the aggregate system and each subsystem. Our approach expands the technology sets of each subsystem by allowing each to acquire resources from the other in exchange for delivery of the appropriate (intermediate or final) product, and to form composites from both subsystems. Managers of each subsystem will not agree to "vertical integration" initiatives unless each subsystem will be more efficient than what each can achieve by separately applying conventional efficiency analysis. A Pareto Efficient frontier characterizes the acceptable set of efficiencies of each subsystem from which the managers will negotiate to select the final outcome. Three proposals for the choice for the Pareto efficient point are discussed: the one that achieves the largest equiproportionate reduction in the classical efficiencies; the one that achieves the largest equal reduction in efficiency; and the one that maximizes the radial contraction in the aggregate consumption of resources originally employed before integration. We show how each choice for the Pareto efficient point determines a derived measure of aggregate efficiency. An extensive numerical example is used to illustrate exactly how the 2 subsystems can significantly improve their operational efficiencies via integration beyond what would be predicted by conventional analysis.

Keywords Multi-stage production systems · Productivity and efficiency measurement · Data envelopment analysis

B. Golany (✉)· U. Passy
Faculty of Industrial Engineering and Management,
Technion—Israel Institute of Technology,
Haifa, 32000, Israel
e-mail: golany@ie.technion.ac.il

S. T. Hackman
School of Industrial and Systems Engineering,
Georgia Institute of Technology,
Atlanta, GA 30332, USA

Introduction

Data Envelopment Analysis (DEA) was developed to measure the relative efficiency of operational units known in the literature as "Decision Making Units"—DMUs (see Charnes et al., 1994). Given a pair of observed input-output vectors (X_0, Y_0), DEA assesses its efficiency by comparing it to other choices in the technology set $\mathcal{T} = \{(X, Y)\}$, which characterizes the collection of all input vectors X that can produce the output vector Y. For example, the classical DEA radial input measure of efficiency (Charnes, Cooper, and Rhodes, 1978) is calculated as

$$\text{Min}\ \{\theta : (\theta X_0, Y_0) \in \mathcal{T}\}. \tag{1}$$

The technology set $\mathcal{T}$ is extrapolated from the observed data on input-output pairs (X_j, Y_j), $j = 1, 2, \ldots, N$, for N DMUs, and is typically defined via a set of linear inequalities, which turns (1) into a linear program. The optimal input-output pair (X^*, Y^*) of the linear program to which (X_0, Y_0) is compared is a linear combination of existing DMUs, and is commonly referred to as a "composite" unit or system.

A significant portion of DEA research to date has focused on defining the rules for constructing $\mathcal{T}$, and defining corresponding measures of efficiency. Regardless of the definition of efficiency, most DEA models treat each DMU as a non-separable entity without attempting to probe the internal mechanisms of how each DMU converts its inputs into outputs. With today's information systems it is now much easier to collect data on how capital and labor are used to transform raw materials through various stages to produce final products. The availability of such data presents an opportunity to explore efficiency measurement of the stages within complex, multi-stage DMUs.

In particular, we are interested in DMUs that consist of several stages arranged in series where succeeding stages (or subsystems) are fed by a mixture of external inputs and intermediate factors which are outputs of preceding stages. The focus of the present paper is on how to assess the efficiency of each stage within the aggregate system and how to explore possible tradeoffs of these efficiencies. As a starting point, of course, one can treat each subsystem as a system in its own right. In this manner, the technology for each subsystem is constructed using the relevant input-output data from its own peers, and the technology of the aggregate system is constructed on the basis of aggregated inputs and outputs and without regard to intermediate input-output factors that link the various stages. As we subsequently demonstrate, this approach exhibits the phenomena in which it is possible for the aggregate system to be rated very inefficient, while each subsystem is rated efficient, and for the aggregate system to be rated near efficient, while each subsystem is rated highly inefficient.

In this paper, we propose an expansion of the ordinary technology sets which were used till now in DEA and develop a corresponding efficiency measurement framework that simultaneously computes the efficiency of each subsystem and the aggregate system. The measurement framework has the following properties. First, if each subsystem is rated efficient, then so must the aggregate system. Second, each subsystem's efficiency and the aggregate efficiency cannot exceed the efficiency obtained using the classical DEA approach, and may be expected to be far lower. Third, the methodology described herein provides the recipe on how to obtain the operational improvements at *all* levels of the hierarchy, as it explicitly integrates the computation of the various efficiencies.

With the assumptions of our expanded technology set, we show how it may be possible to obtain significant improvements in operations when the subsystems agree to "vertically integrate". The key to our approach is that we allow each subsystem to construct composite units from both subsystems as in the classical approach, and we allow each subsystem to *acquire* resources from the other stage in return for delivering that stage's output. We insist that acquisitions by one subsystem from another have to make economic sense in that both parties have to benefit and consent. A Pareto Efficient frontier of acceptable efficiencies for each subsystem will be introduced. While in principle any choice along the Pareto Efficient frontier could be selected, depending on which subsystem has the greater market power, we discuss several reasonable choices. We show how each choice for the Pareto efficient point by the subsystems determines a derived measure of aggregate efficiency. Our choice for the measure of aggregate efficiency corresponds to the Pareto efficient point that achieves the largest radial contraction in the aggregate amount of capital and labor used before integration.

To achieve the benefits listed above, we limit our application to 2 stages in series, and we make the following simplifying assumptions:

1. *Technology.* Each subsystem 1 uses capital and labor to produce an intermediate product used by subsystem 2 to produce final product. Subsystem 2 also requires capital and labor, which are assumed to be completely transferrable resources between stages. All technologies described herein exhibit constant returns-to-scale.
2. *Market.* Each subsystem 2 has a unique supplier given by its subsystem 1. There is a market for the intermediate product, and the transfer price subsystem 1 charges subsystem 2 is the prevailing market price. Competitive markets exist so that either subsystem is a *price-taker* in the input and output markets. That is, it may expand or contract its output without affecting its price or cost of inputs.
3. *Organization.* Monitoring and information costs may make it difficult (and perhaps unwise) for senior management to dictate sweeping changes to the allocation of capital and labor between stages. Organizationally, each subsystem is viewed as a profit center, and each manager is given decision-making authority. Although we do not explicitly model the incentive scheme for the managers, which is beyond the scope of this paper, we assume each manager is highly motivated to improve his own system's efficiency. For our purposes, efficiency may be thought of as the proxy for performance, which is why each manager will not consent to an acquisition of resources unless he directly benefits from it. From an organizational perspective, we view our modeling approach described herein as a natural starting point for efficiency improvement.

Attempts to model DMUs that exhibit known internal structures started in the mid-1980s. Färe and Primont (1984) constructed multi-plant efficiency measures and illustrated their models by analyzing utility firms each of whom operated several electric generation plants. Their structure may be characterized as *horizontal integration* since the plants they model operate in parallel and there is no flow of intermediate inputs or outputs between them. Färe, Grosskopf, and Li (1992) further expand this modeling approach to describe firm and industry performance where, in some cases, reallocation of resources among firms is allowed to improve the industry performance. Cook et al. (1998) define various hierarchies and groupings of DMUs in DEA, and apply DEA formulations for the different groupings.

Another research thrust explored *vertical integration* structures in which a series of stages is connected through intermediate input-output factors. Charnes et al. (1986, see also Charnes et al., 1994, p. 432) developed a two-stage model in the context of the US Army Recruiting Command. The first stage used advertising to generate awareness and propensity to enlist. The two outputs generated by the first stage were joined by other external inputs (e.g. recruiters) to produce the second stage outputs, which were the actual recruitment contracts. Färe and Whittaker (1995) developed a linear programming model that focused on the role played by the intermediate factors and demonstrated its potential through an application to dairy farms.

Hoopes, Triantis, and Partangel (2000) developed a goal-programming DEA formulation that models serial manufacturing processes and applied it to data on circuit board manufacturing. Chen and Zhu (2004) explored two-stage systems in the banking industry (where the first stage produces deposits which are then used to produce loans) and developed a model that identifies the efficiency frontier that characterizes such systems.

Färe and Grosskopf (1996) and (2000) deserve special mention, as they pioneered a line of research, coined *network DEA*, aimed at developing a general multi-stage model with intermediate inputs-outputs. Their representation of the flow of product is consistent with the industrial engineering and operations research literature on multi-stage systems (e.g., Johnson and Montgomery, 1974; Hackman and Leachman, 1989; Troutt, Ambrose, and Chan, 2001; Graves et al., 1993). Each internal stage's technology is modeled using a single-stage DEA model. The conventional radial-based measure of aggregate efficiency would still be determined as in (1) using their more extensive description of technology. For an application of the Färe and Grosskopf framework, see Löthgren and Tambour (2000), who applied their modeling approach to evaluate the performance of Swedish pharmacies.

It is important to point out here that our two-stage model of the *flow of material* is a special case of Färe and Grosskopf's multi-stage framework; however, our proposed aggregate efficiency measure is fundamentally different. In particular, in cases when our proposed aggregate efficiency is higher, it necessarily follows that it would not be possible to disaggregate the Fare-Grosskopf aggregate efficiency measure into separate efficiency measures for which each submanager would consent. If there is an aggregate manager who may unilaterally reallocate resources without consent of the submanagers, then assessing aggregate efficiency using the Färe-Grosskopf framework may lead to superior results for the whole system. However, as we have noted, due to the linkage of inputs and outputs between the stages, in such a context one subsystem's efficiency may be vastly improved at the expense of potential improvement in the other subsystem, which may render meaningless the assessment of subsystem efficiency.

The outline of the paper is as follows. Section 1 introduces the specific network structure and data used throughout the paper, and discusses a motivating example to illustrate how it is possible to achieve significant improvements in efficiency via acquisition. Section 2 presents the formal descriptions of the newly expanded models of technology that are used to assess the efficiencies of each stage, and discusses their application to our dataset. Section 3 introduces the aforementioned Pareto Efficient frontier. Section 4 develops the aggregate measure of efficiency and compares the numerical results obtained from the dataset to both the classical approach and the Färe and Grosskopf approach. Section 5 presents the concept of "consistent pricing" which characterizes both our proposed models and that of Färe and Grosskopf. Section 6 discusses extensions to the basic modeling approach and analyzes, as well as suggestions for further research. Section 7 closes this paper with a few concluding remarks.

1. Preliminaries

1.1. A representative multi-stage system

To ease notational burdens and to make concrete the conceptual discussions to follow, we shall analyze multi-stage systems such as the one depicted in Fig. 1. Each DMU$_j$ ($j =$ 1,..., N) consists of two subsystems in series. Subsystem $1j$ (hereafter abbreviated S_{1j}) uses capital K_{1j} and labor L_{1j} to produce intermediate product I_j. Subsystem $2j$ (hereafter

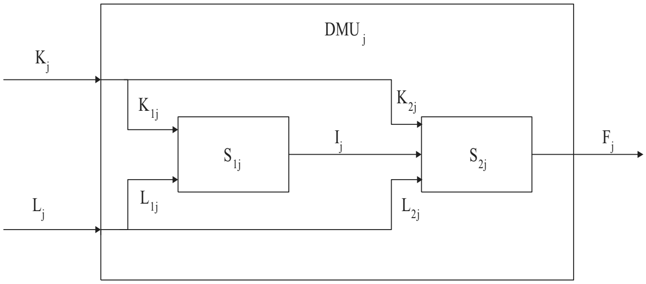

Fig. 1 Aggregate DMU with two stages in tandem

abbreviated by S_{2j}) uses capital K_{2j} and labor L_{2j} together with I_j to produce final output F_j. Constant returns-to-scale (CRS) will be assumed throughout.

The models we develop will be illustrated with a 10 DMU numerical example constructed as follows. We modeled each subsystem's technology via a Cobb-Douglas production function so that the observed output for S_{1j} was $K_{1j}^{\alpha}L_{1j}^{1-\alpha} - 10\mu_1$, and the observed output for S_{2j} was $K_{2j}^{\beta}L_{2j}^{\gamma}I^{1-\beta-\gamma} - 10\mu_2$. The capital and labor inputs were randomly generated in the range [1,100], and we set the other parameters as $\alpha = 0.4$, $\beta = 0.25$, $\gamma = 0.45$ and $\mu_1, \mu_2 \sim$ Uniform[0,1]. The resulting data on inputs and outputs for S_{1j}, (K_{1j}, L_{1j}, I_j), and S_{2j}, $(K_{2j}, L_{2j}, I_j, F_j)$, are given in Table 1.

1.2. Classical models of technology

Following Shephard (1970); Charnes, Cooper, and Rhodes (1978); Banker, Charnes, and Cooper (1984); Färe and Grosskopf (1996), a model of technology $\mathcal{T} \in R_+^n \times R_+^m$ characterizes the collection of all input vectors $X \in R_+^n$ that can be used to produce the output vector $Y \in R_+^m$. For a given dataset of input-output pairs (X_j, Y_j), $j = 1, 2, \ldots, N$, the classical DEA approach models the technology as

$$\mathcal{T} \equiv \left\{(X, Y) : \sum_j \lambda_j X_j \leq X, \sum_j \lambda_j Y_j \geq Y\right\}, \tag{2}$$

Table 1 Data for the numerical example

DMU	K_1	K_2	K	L_1	L_2	L	I	F
1	32	81	113	67	83	150	46.928	64.941
2	96	28	124	40	81	121	47.431	46.492
3	79	51	130	89	79	168	79.694	67.388
4	41	80	121	35	26	61	32.978	31.124
5	99	8	107	33	74	107	45.921	35.018
6	72	29	101	15	36	51	24.861	29.146
7	21	88	109	64	23	87	32.250	34.049
8	60	39	99	71	49	120	64.659	45.176
9	7	86	93	80	16	96	21.531	21.062
10	10	40	50	33	11	44	12.519	10.189

Table 2 Classical efficiency evaluation for S_{1j} and S_{j2}

	S_{1j}		S_{2j}	
DMU_j	θ_{1j}^{CL}	Benchmarks	θ_{2j}^{CL}	Benchmarks
1	1.00	1	1.00	1
2	0.943	5,8	1.00	2
3	0.973	5,8	1.00	3
4	0.958	5,8	0.904	3,6
5	1.00	5	1.00	5
6	1.00	6	1.00	6
7	0.941	1,9	1.00	7
8	1.00	8	1.00	8
9	1.00	9	0.918	1,7
10	0.748	1,9	0.735	1,7

where we now and hereafter suppress the nonnegativity constraints imposed on the intensity variables (the λ_j's). Given the technology $\mathcal{T}$ and using the subscript "0" to denote a DMU in our dataset which is to be analyzed (i.e., $(X_0, Y_0) \in \mathcal{T}$), the classical (CL) radial measure of input efficiency is defined as

$$\theta_0^{CL} \equiv \text{Min}\,\{\theta_0 : \ (\theta_0 X_0, Y_0\} \in \mathcal{T}\}. \tag{3}$$

(In principle, any measure of input efficiency would suffice for the developments to follow, since our focus is to expand the classical model of technology when detailed information about its structure is available). For the multi-stage systems depicted in Fig. 1, the classical descriptions of technology for each subsystem are

$$\mathcal{T}_1 \equiv \left\{((K, L),\ I) : \sum_j \lambda_{1j} K_{1j} \le K,\ \sum_j \lambda_{1j} L_{1j} \le L,\ \sum_j \lambda_{1j} I_j \ge I\right\}, \tag{4}$$

$$\mathcal{T}_2 \equiv \left\{((K, L, I),\ F) : \sum_j \lambda_{2j} K_{2j} \le K,\ \sum_j \lambda_{2j} L_{2j} \le L,\ \sum_j \lambda_{2j} I_j \le I,\ \sum_j \lambda_{2j} F_j \ge F\right\}. \tag{5}$$

For each DMU_0 in our dataset the classical measures of input efficiency for each stage are computed as follows:

$$\theta_{10}^{CL} \equiv \text{Min}\{\theta_{10} : \ ((\theta_{10} K_{10}, \theta_{10} L_{10}),\ I_0) \in \mathcal{T}_1\}, \tag{6}$$

$$\theta_{20}^{CL} \equiv \text{Min}\{\theta_{20} : \ ((\theta_{20} K_{20}, \theta_{20} L_{20}, \theta_2 I_0),\ F_0) \in \mathcal{T}_2\}. \tag{7}$$

Computational results are reported in Table 2.

1.3. An expanded model of technology: A motivating example

A numerical example using one of the DMUs in our dataset will be used to explain how to expand the technology to provide better opportunities for each subsystem to improve its efficiency. The model used to generate this example is formally described in the next section. In what follows, we have made the following assumptions: (1) Each stage is managed as a profit center; (2) S_{1j} may sell its intermediate product on the open market for the same price it charges S_{2j}; and (3) Both S_{1j} and S_{2j} may sell any amount of their respective outputs on the open market without affecting input cost or price.

The observed S_{25} of DMU_5 uses 8 units of capital, 74 units of labor and 45.92 units of intermediate product to produce 35.02 units of final product. The manager of S_{25} (hereafter named 'M_{25}'), while always looking to improve efficiency, is content for now as his system is rated efficient by classical efficiency analysis. Now suppose the manager of S_{15} (hereafter named 'M_{15}') comes to M_{25} with the following proposal: "I can show you how to increase your output by 6.3%, while *simultaneously* reducing your cost of inputs by 11.75%. Interested?" That is, M_{15} is proposing a way for M_{25} to use $(K, L, I) = (7.06, 65.29, 40.52)$ to produce $F = 37.23$ instead of M_{25}'s current production plan that uses $(K, L, I) = (8, 74, 45.92)$ to produce $F = 35.02$. To expand his output by 6.3%, M_{25} would normally expect (under CRS) to have to increase his inputs by 6.3%, and so, in effect, M_{15} is offering M_{25} to consume only $100(1 - 0.1175)/1.063 = 83\%$ of his input to achieve the same output level. While M_{25} is obviously intrigued by M_{15}'s proposal, M_{25} demands an explanation as to how M_{15} proposes to accomplish this seemingly impossible task, as M_{25} knows that *both* S_{15} and S_{25} were rated input efficient by classical analysis. M_{15} obliges with the following explanation.

Using classical descriptions of technology for each subsystem, M_{15} found a composite subsystem 1 process that uses (34.48, 40.80) units of capital and labor to produce 37.16 units of intermediate product, and a composite subsystem 2 process that uses (37.04, 45.98, 31.76) units of capital, labor and intermediate product to produce the 37.23 units of final product, which M_{15} promised to deliver to M_{25}. The total amounts of capital and labor required by these two composite processes are 71.52 and 86.78, respectively. With the 7.06 units of capital and 65.29 units of labor acquired from M_{25}, M_{15} still needs 64.46 units of capital and 21.49 units of labor, which he possesses as these totals represent only 65.1% of his current capacity of (99, 33) units of capital and labor. With respect to the intermediate product, M_{25} notes that while he is now purchasing $45.92-40.52 = 5.40$ *less* units of intermediate product from M_{15}, the difference between what the subsystem 2 composite requires and what the subsystem 1 composite currently produces of intermediate product is also $5.40 = 37.16 - 31.76$ units, which M_{15} will sell on the open market to compensate him for the loss in revenue from M_{25}. M_{25} is satisfied that M_{15}'s proposal is conceptually sound.

M_{25} now understands why M_{15} is so eager to offer this proposal to M_{25}: under the proposal, M_{15} will be able to free up 34.9% of *his* inputs, a considerable savings, while still producing his same level of output. Since M_{15} cannot achieve this savings without M_{25}'s consent, M_{25} realizes he must understand exactly how M_{15} was able to devise this seemingly ingenious plan, so that he will be in position to negotiate with M_{15} a better deal for himself.

2. The expanded technology sets for S_{1j} and S_{2j}

For every DMU_j, the manager of S_{1j} now realizes that the classical efficiency analysis constructed the efficient frontier using *only* subsystem 1 processes. It does not consider the possibility that M_{1j} may have the options of adopting an alternative subsystem 2 production process *and* acquiring resources from M_{2j} (as long as M_{2j} would agree). With these options, the technology set for S_{1j}, which defines the collection of input pairs (K, L) that can produce at least I_j, has been expanded.

Under the CRS assumption, M_{10} knows that $\omega_{20}((K_{20}, L_{20}, I_0),\ F_0) \in \mathcal{T}_2$ for all $\omega_{20} \geq 0$. In order to entice M_{20} to agree, M_{10} selects a value $\theta_{20} < 1$, and offers M_{20} the opportunity to achieve the input-output point of $\omega_{20}((\theta_{20}K_{20}, \theta_{20}L_{20}, \theta_{20}I_0),\ F_0)$. In order for M_{10} to meet his obligation to M_{20} and his objective, namely to produce I_0 with resources (K, L), he must find two composite processes $((\hat{K}_1, \hat{L}_1),\ \hat{I}_1) \in \mathcal{T}_1$ and $((\hat{K}_2, \hat{L}_2, \hat{I}_2),\ \hat{F}) \in \mathcal{T}_2$ for which the following 4 *inventory balance equations* must hold:

[E1] *Capital.* The "supply" of capital from M_{10} and M_{20}, $K + \omega_{20}(\theta_{20}K_{20})$, must be no smaller than the "demand" for capital by both composite subsystems, $\hat{K}_1 + \hat{K}_2$;
[E2] *Labor.* The "supply" of labor from M_{10} and M_{20}, $L + \omega_{20}(\theta_{20}L_{20})$, must be no smaller than the "demand" for labor by both composite subsystems, $\hat{L}_1 + \hat{L}_2$;
[E3] *Intermediate Product.* The "supply" of intermediate product from M_{20} and the composite Stage 1 process, $\hat{I}_1 + \omega_{20}(\theta_{20}I_0)$, must be no smaller than the "demand" for intermediate product by M_{10} and the composite Stage 2 process, $I_0 + \hat{I}_2$; and
[E4] *Final Product.* The "supply" of final product from the composite Stage 2 process, $\hat{F}$, must be no smaller than the "demand" for final product by M_{20}, $\omega_{20}F_0$.

Let $\mathcal{T}_1^E(\theta_{20})$ denote the collection of input-output pairs $((K, L),\ I_0)$ that satisfy the inventory balance equations [E1—E4] listed above for DMU$_0$. Given θ_{20}, it would make sense for M_{10} to find the least amount of capital and labor to satisfy his own output requirement of I_0. Accordingly, he should solve the following linear programming model, which we shall denote as the *Acquisition* (AQ) model:

$$\theta_{10}^{AQ}(\theta_{20}) \equiv \min \theta_{10}$$

$$\sum_j \lambda_{1j}K_{1j} + \sum_j \lambda_{2j}K_{2j} \le \theta_{10}K_{10} + \omega_{20}[\theta_{20}K_{20}]$$

$$\sum_j \lambda_{1j}L_{1j} + \sum_j \lambda_{2j}L_{2j} \le \theta_{10}L_{10} + \omega_{20}[\theta_{20}L_{20}]$$

$$\sum \lambda_{1j}I_j + \omega_{20}[\theta_{20}I_0] \ge I_0 + \sum \lambda_{2j}I_j$$

$$\sum \lambda_{2j}F_j \ge \omega_{20}F_0 \tag{8}$$

In the proposal of M_{15} to M_{25} that was described in Section 1.3, M_{15} selected $\theta_{25} = 0.83$, and solved the AQ model, whose solution was $\omega_{25} = 1.063$ with $\theta_{15}^{AQ}(\theta_{25}) = 0.651$.

Now that M_{20} understands how M_{10} was able to achieve *his* objective, M_{20} realizes he can play the same game. Let $\mathcal{T}_2^E(\theta_{10})$ denote the collection of input-output pairs $((K, L, I),\ F_0)$ that satisfy analogous 4 inventory balance requirements as described above. Given θ_{10} it would make sense for M_{20} to find the least amount of capital and labor to satisfy his own output requirement of F_0. Accordingly, he would solve his own *Acquisition* (AQ) model, namely, the following linear programming model:

$$\theta_{20}^{AQ}(\theta_{10}) \equiv \min \theta_{20}$$

$$\sum_j \lambda_{1j}K_{1j} + \sum_j \lambda_{2j}K_{2j} \le \theta_{20}K_{20} + \omega_{10}[\theta_{10}K_{10}]$$

$$\sum_j \lambda_{1j}L_{1j} + \sum_j \lambda_{2j}L_{2j} \le \theta_{20}L_{20} + \omega_{10}[\theta_{10}L_{10}]$$

$$\sum \lambda_{1j}I_j + \theta_{20}I_0 \ge \sum \lambda_{2j}I_j + \omega_{10}I_0$$

$$\sum \lambda_{2j}F_j \ge F_0 \tag{9}$$

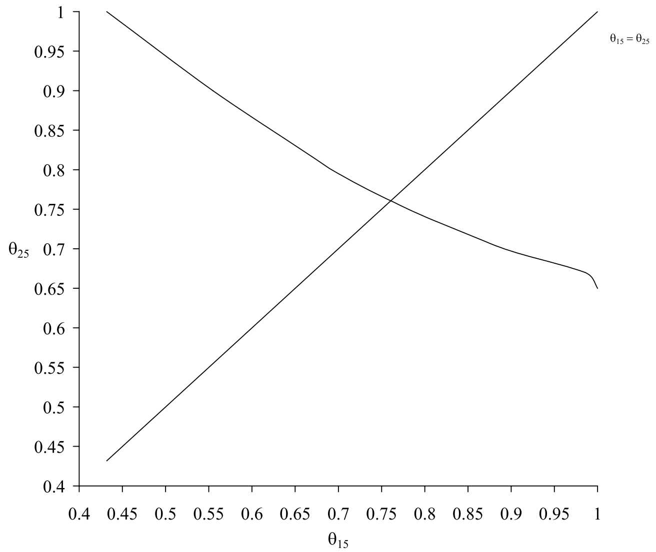

Fig. 2 Efficient frontier for DMU_5, θ_{15} *vs.* θ_{25}

For example, suppose M_{25} selects $\theta_{15} = 0.9$. Solution of his AQ model gives $\omega_{15} = 0.686$ and $\theta_{25}^{AQ}(\theta_{15}) = 0.697$. Note how much better off M_{25} is and worse off M_{15} is as compared to M_{15}'s original proposal. Both managers will agree that either proposal will outperform the classical analysis.

We close this section by emphasizing the following point about describing the subsystem technologies. Since we allow the possibility of one subsystem manager to acquire resources from the other, as long as they can agree, the potential acquisition of resources consistent with the "θ_{10}—θ_{20}" agreement must now be embedded in the respective descriptions of technology given by $\mathcal{T}_1^E(\theta_{20})$ and $\mathcal{T}_2^E(\theta_{10})$ to reflect the set of all production possibilities.

3. Pareto efficient frontiers

It should be intuitively clear that for the serial system we discuss here a gain by one manager is a loss by the other manager. Regardless of the final choice for how the two subsystems shall vertically integrate, the agreed-upon choice for θ_{10} and θ_{20} should minimally result in a Pareto efficient outcome; that is, $(\theta_{10}, \theta_{20}) = (\theta_{10}^{AQ}(\theta_{20}), \theta_{20}^{AQ}(\theta_{10}))$. Otherwise, neither manager, M_{10} nor M_{20}, would agree to the vertical integration.

The efficient frontier corresponding to DMU_5 in our example is depicted in Fig. 2. This frontier was constructed using a recently developed algorithm by Hackman and Passy (2002). When θ_{15} is set to 1.00, θ_{25} is at its lowest value 0.66. On the other hand, when θ_{25} is set to 1.00, θ_{15} is assigned its lowest value 0.43. We remark that the frontier does not always span an interval as depicted in Fig. 2. In extreme cases the frontier may consist of a single point. Such cases are similar, in a sense, to the class of "weakly efficient" DMUs that were discussed in Charnes, Cooper, and Thrall (1986).

Two remarks concerning the Acquisition Models (8) and (9) that determine the Pareto Efficient frontier are in order. First, it is not necessary to solve both Acquisition Models, as there is a one-to-one correspondence between the solutions for each Acquisition Model: the solution to Model (9) may be obtained from the solution to Model (8) by dividing λ^*_{1j} and λ^*_{2j} by ω^*_{20}, and setting $\omega^*_{10} = (\omega^*_{20})^{-1}$. Second, the solutions to either Model (8) or Model (9) must necessarily lie below their respective classical efficiency counterparts: the linear program to compute θ^{CL}_{10} is a special case of Model (8) in which $\omega_{20} = 0$ and $\lambda_{2j} = 0$, and the linear program to compute θ^{CL}_{20} is a special case of Model (9) in which $\omega_{10} = 0$ and $\lambda_{1j} = 0$. From an economic perspective, M_{20} would never agree to a proposal from M_{10} if the proposed θ_{20} exceeds what he could achieve on his own, and M_{10} would never offer a proposal to M_{20} in which he receives an efficiency θ_{10} that exceeds what he could achieve on his own, too.

In principle, any point on the Pareto efficient frontier is a candidate. One natural choice is to select the point that achieves the largest *equiproportionate* reduction in the respective classical single-stage efficiencies θ^{CL}_{10} and θ^{CL}_{20}. For example, the 45° line in Fig. 2 intersects the frontier at the equiproportional point $\theta_{15} = \theta_{25} = 0.762$, a point which might be considered as "fair" for both subsystems. A second choice is to select the point that achieves the largest equal reduction in efficiency. (When $\theta^{CL}_{10} = \theta^{CL}_{20} = 1$, as is the case for DMU_5, these two choices will obviously coincide). A third choice is to select the point that achieves for the vertically integrated unit the largest radial contraction in the aggregate amounts of capital and labor originally employed, which we more fully discuss in the next section.

4. Aggregate efficiency

4.1. Measures of aggregate input efficiency

From the perspective of an aggregate DMU, the classical model of technology is given by

$$\mathcal{T}_A = \left\{((K, L),\ F) :\ \sum_j \lambda_j (K_{1j} + K_{2j}) \le K, \right. \tag{10}$$

$$\left. \sum_j \lambda_j (L_{1j} + L_{2j}) \le L,\ \sum_j \lambda_j F_j \ge F \right\}.$$

For the aggregate DMU_0 (denoted hereafter as $A0$) the classical model ignores the intermediate product I_0 as it represents internal production. The corresponding classical input efficiency measure would be computed as:

$$\theta^{CL}_{A0} = \text{Min}\{\theta_{A0} :\ ((\theta_{A0}(K_{10} + K_{20}), \theta_{A0}(L_{10} + L_{20})),\ F_0) \in \mathcal{T}_A\}. \tag{11}$$

Färe and Grosskopf (1996) provide an in-depth development of models of technology for general multi-stage systems. One of their basic models (Färe and Grosskopf, 1996, pp. 20–23) allows *complete transferability* (CT) of capital and labor flows between the stages. Applied to the two-stage systems we analyze in this paper, their model is formulated as:

$$\mathcal{T}^{CT}_A = \left\{((K, L),\ F) :\ \sum_j \lambda_{1j} K_{1j} + \sum_j \lambda_{2j} K_{2j} \le K, \right. \tag{12}$$

$$\sum_j \lambda_{1j} L_{1j} + \sum_j \lambda_{2j} L_{2j} \le L, \tag{13}$$

$$\sum_j \lambda_{1j} I_j \;-\; \sum_j \lambda_{2j} I_j \geq 0 \tag{14}$$

$$\left.\sum_j \lambda_{2j} F_j \geq F\right\}. \tag{15}$$

The third constraint above represents *inventory balance* of intermediate product to ensure that the supply of I produced by the composite S_1 will be sufficient to satisfy the demand for I by the composite S_2. Assuming complete transferability of resources, the Färe-Grosskopf measure of input efficiency would be computed as:

$$\theta_{A0}^{CT} = \text{Min}\left\{\theta_{A0} :\ ((\theta_{A0}(K_{10} + K_{20}), \theta_{A0}(L_{10} + L_{20})),\ F_0) \in \mathcal{T}_A^{CT}\right\}. \tag{16}$$

4.2. A derived measure of aggregate efficiency

For each Pareto efficient point $(\theta_{10}, \theta_{20})$, let $K(\theta_{10}, \theta_{20})$ and $L(\theta_{10}, \theta_{20})$ denote, respectively, the aggregate amounts of capital and labor which the vertically integrated unit would use to produce *both* F_0 and I_0. A natural choice for a derived measure of aggregate input efficiency is

$$\theta_{A0}^{D}(\theta_{10}, \theta_{20}) \;\equiv\; \text{Max}\left\{\frac{K(\theta_{10}, \theta_{20})}{K_{10} + K_{20}},\ \frac{L(\theta_{10}, \theta_{20})}{L_{10} + L_{20}}\right\}. \tag{17}$$

We now show how to compute $K(\theta_{10}, \theta_{20})$ using both Acquisition Models (8) and (9). (The derivation for $L(\theta_{10}, \theta_{20})$ is analogous). First suppose that $\omega_{20} \leq 1$. Examine the right-hand side of the first constraint in (8). In return for delivering $\omega_{20} F_0$ units of final product to S_{20} and meeting its own requirements of producing I_0, S_{10} uses $\omega_{20}[\theta_{20} K_{20}]$ units of capital it acquires from S_{20} and $\theta_{10} K_{10}$ for its own production needs. For the vertically integrated unit to produce a total of F_0, S_{20} will have to produce the remaining amount $(1 - \omega_{20}) F_0$ by its own production process and thereby consume $(1 - \omega_{20}) K_{20}$ units of capital. In this case $K(\theta_{10}, \theta_{20}) = \theta_{10} K_{10} + \omega_{20}[\theta_{20} K_{20}] + (1 - \omega_{20}) K_{20}$. Now suppose $\omega_{20} \geq 1$. Since $\omega_{10} = \omega_{20}^{-1} \leq 1$, we shall examine the right-hand side of the first constraint in (9). Here, in return for delivering $\omega_{10} I_0$ units of intermediate output to S_{10}, S_{20} uses $\omega_{10}[\theta_{10} K_{10}]$ of capital it acquires from S_{10}, and $\theta_{20} K_{20}$ units for its own production needs. For S_{10} to produce a total of I_0, it will need $(1 - \omega_{10}) K_{10}$ units of capital to produce the remaining amount $(1 - \omega_{10}) I_0$ using its current production process. In this case $K(\theta_{10}, \theta_{20}) = \omega_{10}[\theta_{10} K_{10}] + \theta_{20} K_{20} + (1 - \omega_{10}) K_{10}$. To summarize, we have

$$K(\theta_{10}, \theta_{20}) \equiv \begin{cases} \theta_{10} K_{10} \;+\; \omega_{20}[\theta_{20} K_{20}] + (1 - \omega_{20}) K_{20}, & \omega_{20} \leq 1 \\ \omega_{10}[\theta_{10} K_{10}] \;+\; \theta_{20} K_{20} + (1 - \omega_{10}) K_{10}, & \omega_{10} \leq 1 \end{cases} \tag{18}$$

$$L(\theta_{10}, \theta_{20}) \equiv \begin{cases} \theta_{10} L_{10} \;+\; \omega_{20}[\theta_{20} L_{20}] + (1 - \omega_{20}) L_{20}, & \omega_{20} \leq 1 \\ \omega_{10}[\theta_{10} L_{10}] \;+\; \theta_{20} L_{20} + (1 - \omega_{10}) L_{10}, & \omega_{10} \leq 1 \end{cases} \tag{19}$$

A numerical example will help explain our proposed derived measure of aggregate efficiency. We solved Model (8) for DMU_3 with $\theta_{23} = 0.9$. The result is $\theta_{13} =$

0.9667 and $\omega_{23} = 0.1439$. The two composite subsystems constructed by the linear program are, respectively, $((\hat{K}_1, \hat{L}_1),\ \hat{I}_1) = ((70.88, 83.87),\ 76.38)$ and $((\hat{K}_2, \hat{L}_2), \hat{I}_2),\ \hat{F}_2) = ((12.09, 12.39, 7.01),\ 9.696)$. Now consider the 4 inventory balance eq. [E1—E4] associated with this Pareto efficient point ($\theta_{13} = 0.9667, \theta_{23} = 0.9$):

$$70.88 + 12.09 \leq (0.9667)[79] + 0.1439[0.90 \cdot 51] = 82.97 \tag{20}$$

$$83.87 + 12.39 \leq (0.9667)[89] + 0.1439[0.90 \cdot 79] = 96.26 \tag{21}$$

$$76.38 + 0.1439[0.90 \cdot 79.694] \geq [79.694] + 7.01 \tag{22}$$

$$9.696 \geq 0.1439(67.39) \tag{23}$$

Note how S_{13} is only promising to deliver 14.39% of final output; the remaining 85.61% must be produced by S_{23} using its own production process. The derived capital in this case is $(0.9667)[79] + 0.1439[0.90 \cdot 51] + (1 - 0.1439) \cdot 51 = 126.63$, and the derived labor is $(0.9667)[89] + 0.1439[0.90 \cdot 79] + (1 - 0.1439) \cdot 79 = 163.89$. When (126.63, 163.89) is compared to the original values of (130, 168), we obtain $\theta^D_{A3} = 0.9755$.

The derived aggregate measure of efficiency is measured along the Pareto efficient frontier corresponding to Models (8) and (9). It can never be larger than 1.0. Conceptually, any point on the Pareto efficient frontier could be used to define the aggregate efficiency. As discussed at the end of the last section there are two obvious choices: the equiproportional solution, $(\theta_{10} = \rho\theta^{CL}_{10}, \theta_{20} = \rho\theta^{CL}_{20})$, where $\rho \leq 1$, and the equal contraction solution in which $\theta_{10} = \theta_{20}$. We propose a third alternative: Minimize θ^D_{A0} on the Pareto efficient frontier, which we shall denote by θ^P_{A0}. To compute θ^P_{A0}, a bi-level programming problem, we iteratively solve Model (9) (resp. Model (8)) for different θ_{10} (resp. θ_{20}) values.

4.3. Computational results

Table 3 reports the computational results for each measure of aggregate efficiency. First, we compare θ^{CT}_{Aj} to θ^{CL}_{Aj} for $j = 1, \ldots, 10$. In stark contrast to the relative efficiency nature of DEA, when additional flexibility of transferring resources between stages is available, the CT model is able to use this flexibility to identify potential improvement opportunities for *all* DMUs. Indeed, the relevant benchmarks, which report the reference sets for each of the evaluated DMUs, contain *both* S_{1j} and S_{2j} stages (first and second rectangular brackets, respectively, in column 5 of Table 3). Of course, from a measurement perspective it will always be the case that $\theta^{CT}_{Aj} \leq \theta^{CL}_{Aj}$.

When comparing θ^{CT}_{Aj} to θ^P_{Aj} in Table 3, we see that $\theta^P_{Aj} > \theta^{CT}_{Aj}$ always holds. (We have been unable to establish any definitive relationship between θ^{CL}_{Aj} and θ^P_{Aj}). Thus, for our numerical example, the CT model indeed finds the maximal possible contraction from the point of view of the aggregate DMU. From an organizational perspective, it may not be possible to implement this solution. We know it is impossible to achieve a better result than θ^P_{Aj} *along the Pareto Efficient frontier*. Consequently, to implement the solution proposed by θ^{CT}_{Aj} when it is smaller than θ^P_{Aj} will require either M_{1j} or M_{2j} to consent to a restructuring that would make him *worse* off than he can achieve by negotiating directly with the manager of the other subsystem. When consent is required, it would make more sense for an aggregate manager to select a Pareto efficient point that both managers will accept. Table 4 records the Pareto efficient points for the two stages and their corresponding θ^D_{Aj} values. For every DMU_j, the minimal value for the derived aggregate efficiency (θ^P_{Aj}) is given in a box and the equiproportional choice is highlighted in boldface. Observe the wide disparity in efficiencies

Table 3 Aggregate efficiency measures

	Classical efficiency		Complete transferability		Expanded technology	
DMU_j	θ_{Aj}^{CL}	Benchmarks	θ_{Aj}^{CT}	Benchmarks	θ_{Aj}^{P}	Benchmarks
1	1.00	1	0.983	[1,8], [6]	0.988	[8], [6]
2	0.829	1,6	0.777	[5,8] [1]	0.842	[8] [1]
3	0.922	1,6	0.899	[1,8], [6]	0.953	[1], [6]
4	0.893	6	0.853	[5], [1,7]	0.922	[5], [1]
5	0.710	1,6	0.666	[5,8], [1]	0.781	[8], [6]
6	1.00	6	0.956	[5], [1,7]	0.981	[5], [7]
7	0.799	1,6	0.751	[5,8], [1]	0.867	[8], [6]
8	0.854	1,6	0.816	[8], [1,6]	0.868	[-], [6]
9	0.480	1,6	0.449	[5,8], [1]	0.621	[8], [6]
10	0.486	1,6	0.456	[5,8], [1]	0.712	[8], [6]

Table 4 Derived aggregate efficiency along the Pareto frontier of the two stages

DMU	Measures	Efficiency values								
1	θ_{11}	1.000	0.9999	**0.9865**	0.9703	0.9272				
	θ_{21}	0.9837	0.9837	**0.9865**	0.9900	1.000				
	θ_{A1}^{D}	0.991	0.990	0.989	0.988	0.990				
2	θ_{12}	0.9428	0.9409	0.9344	**0.8358**	0.8037	0.7638	0.7427	0.7014	0.5196
	θ_{22}	0.7970	0.7980	0.800	**0.8358**	0.8500	0.8600	0.8800	0.9000	1.000
	θ_{A2}^{D}	0.924	0.923	0.919	0.861	0.845	0.842	0.844	0.867	0.892
3	θ_{13}	0.9731	0.9718	0.9667	**0.9288**	0.8408	0.766	0.5266		
	θ_{23}	0.8753	0.8800	0.9000	**0.9288**	0.9500	0.9600	1.000		
	θ_{A3}^{D}	0.986	0.981	0.976	0.953	0.958	0.961	0.965		
4	θ_{14}	0.9580	0.9436	0.9292	**0.8936**	0.5912				
	θ_{24}	0.8600	0.8700	0.8800	**0.8936**	0.900				
	θ_{A4}^{D}	0.986	0.947	0.922	0.952	0.952				
5	θ_{15}	1.000	0.8924	**0.7608**	0.6922	0.6511	0.5547	0.4318		
	θ_{25}	0.6600	0.7000	**0.7608**	0.8000	0.8300	0.9000	1.000		
	θ_{A5}^{D}	1.000	0.908	0.819	0.781	0.781	0.843	0.892		
6	θ_{16}	1.000	0.9827	**0.9612**	0.9523	0.7469	0.6419			
	θ_{26}	0.9000	0.9300	**0.9612**	0.9750	0.9900	1.000			
	θ_{A6}^{D}	1.000	0.981	0.981	0.981	0.982	0.988			
7	θ_{17}	0.9416	0.8728	**0.7666**	0.7367	0.6962	0.5793			
	θ_{27}	0.4917	0.6250	**0.7666**	0.800	0.85	1.000			
	θ_{A7}^{D}	0.988	0.930	0.867	0.879	0.888	0.985			
8	θ_{18}	0.9976	0.9055	**0.8401**	0.6560	0.4822	0.1625			
	θ_{28}	0.8170	0.8300	**0.8401**	0.8740	0.9155	1.000			
	θ_{A8}^{D}	0.927	0.919	0.913	0.892	0.868	0.907			
9	θ_{19}	1.000	0.8086	0.6005	**0.5424**	0.4719	0.3966	0.1618		
	θ_{29}	0.3402	0.400	0.500	**0.5424**	0.600	0.700	0.9065		
	θ_{A9}^{D}	0.825	0.771	0.687	0.621	0.723	0.813	0.913		
10	$\theta_{1,10}$	0.7476	0.7000	**0.5589**	0.5226	0.4896	0.4386			
	$\theta_{2,10}$	0.3351	0.4000	**0.5589**	0.65	0.6670	0.6967			
	$\theta_{A,10}^{D}$	0.949	0.821	0.712	0.726	0.763	0.809			

for each stage corresponding to the Pareto aggregate efficiency. Since the equiproportionate choice seems to sacrifice little in the way of aggregate efficiency, it may be a practical alternative that is easier for the managers to agree on.

5. The consistent pricing principle

The dual linear fractional program (known in the DEA literature as the *multiplier* formulation) to each manager's Acquisition Model provides an alternative means to understand the tradeoff inherent in the Pareto Efficient frontier for managers M_{10} and M_{20}. For M_{10} we have

$$\theta_{10}^{AQ} = \max \frac{\pi_I I_0}{\pi_K K_{10} \; + \; \pi_L L_{10}}$$

$$\left.\begin{array}{r} \dfrac{\pi_I I_j}{\pi_K K_{1j} + \pi_L L_{1j}} \leq 1 \\ \dfrac{\pi_F F_j}{\pi_K K_{2j} + \pi_L L_{2j} + \pi_I I_j} \leq 1 \end{array}\right\} j = 1, \ldots, n$$

$$\frac{\pi_F F_0}{\pi_K K_{20} + \pi_L L_{20} + \pi_I I_0} \geq \theta_{20}, \tag{24}$$

and for M_{20} we have

$$\theta_{20}^{AQ} = \max \frac{\pi_F F_0}{\pi_K K_{20} \; + \; \pi_L L_{20} \; + \; \pi_I I_0}$$

$$\left.\begin{array}{r} \dfrac{\pi_I I_j}{\pi_K K_{1j} + \pi_L L_{1j}} \leq 1 \\ \dfrac{\pi_F F_j}{\pi_K K_{2j} + \pi_L L_{2j} + \pi_I I_j} \leq 1 \end{array}\right\} j = 1, \ldots, n$$

$$\frac{\pi_I I_0}{\pi_K K_{10} + \pi_L L_{10}} \geq \theta_{10}. \tag{25}$$

The last constraint in each model ensures a lower bound on the efficiency of the counterpart subsystem. As the lower bound parameter varies it changes the efficiency in the obvious way: for example, raising θ_{20} lowers S_{10}'s efficiency in (24), and raising θ_{10} lowers S_{20}'s efficiency in (25).

Observe that there is a single multiplier π_K for both K_1 and K_2, a single multiplier π_L for both L_1 and L_2, and a single multiplier π_I that is used to weigh the intermediate factor both when it is an output (of the first stage) and when it is an input (to the second stage). Since the capital, labor and intermediate product are freely transferable between stages, their respective weights in the multiplier formulation should be the same. We shall call this the *Consistent Pricing Principle*. Consistent pricing holds for the Färe-Grosskopf model as well. There, the linear fractional programming dual is given by:

$$\theta_{20}^{CT} = \max \frac{\pi_F F_0}{\pi_K (K_{10} + K_{20}) \; + \; \pi_L (L_{10} + L_{20}) \; + \; \pi_I I_0}$$

$$\left.\begin{array}{l}\dfrac{\pi_I I_j}{\pi_K K_{1j} + \pi_L L_{1j}} \leq 1 \\ \dfrac{\pi_F F_j}{\pi_K K_{2j} + \pi_L L_{2j} + \pi_I I_j} \leq 1\end{array}\right\} j = 1, \ldots, n \tag{26}$$

In an ordinary application of DEA, M_{10} would prefer a larger value of the multiplier π_I, whereas M_{20} would prefer a smaller value of π_I. When both output-input ratios appear in the same optimization, necessarily there will be a tradeoff between the measurement of efficiency of both stages. The consistent pricing principle leads to a natural conflict between the efficiency measures of the two stages. A weighting scheme that might make M_{10} efficient might very well make M_{20} look inefficient, and vice-versa. Thus, there will be a need to coordinate the choice for this multiplier. Regardless of the weighting scheme ultimately agreed upon, it should not be possible to select an alternative set of weights that would make both stages at least as efficient while making one of them more efficient. That is, it should minimally result in a Pareto efficient outcome; otherwise, neither manager M_{10} nor M_{20} would agree to the vertical integration.

6. Extensions and directions for further research

Our aim in this paper was to present a novel approach that opens new ways to evaluate the efficiency of DMUs that are composed of subsystems arranged in series. Due to the complexity of the various relationships we model, we left many possible extensions to the basic model and to the analyzes for further research. These extensions are outlined below.

6.1. Extensions to the basic model

Technology: Our basic model is based on the CCR model that generates the standard radial measure of efficiency. It may be easily extended to the BCC model (Banker, Charnes, and Cooper, 1984) or, more generally, to other technologies used to generate "Russell-type" measures of efficiency that eliminate the slacks in resource use (Russell, 1985).

Structure: Our basic model contains just two stages. An obvious extension is to increase the number of serial stages in the model. Such an extension is possible as the notion of Pareto efficient frontier and the definition of Pareto aggregate efficiency easily generalize. The main (and significant) difficulty here would be the time it would take to compute the various measures.

Choice of variables: The model may also be extended to allow more flexibility in the definition of the inputs and outputs. First, it is straightforward to incorporate additional input and output factors. Second, the model can be extended to allow the presence of inputs and outputs that are not completely transferrable in some subsystems. For example, some inputs may be specific to a particular subsystem and can not be shared. Then, for the purpose of modeling and computing a subsystem's efficiency, it may be easier to work with the multiplier formulation using the consistent pricing principle.

Transaction costs: The basic model assumes no transaction costs when resources are moved between the two stages. In real-world cases, such movements are often accompanied by some transaction costs. Formulating these costs can be done either by adding some terms to the balance equation of the capital or by applying a certain "depreciation" term on each amount that is transferred.

6.2. Extensions to the analysis

Principal-agent issues: We have assumed that subsystem managers have full decision-making power on how to reallocate capital and labor to improve their respective operations. An agreed-upon Pareto efficient point of subsystem efficiencies may or may not lead to the best improvement in aggregate efficiency, which can only be achieved if there is a single "aggregate" manager who has the full authority to unilaterally make decisions. A potentially important and fruitful line of research would be to explicitly define the relationship between the aggregate and subsystem managers using a principal-agent framework (see Jehle and Reny, 2001 for a discussion). That is, do compensation schemes exist that will provide the necessary incentives to subsystem managers to choose the Pareto efficient point desired by the aggregate manager? Typically, such investigations also assume that the principle (the aggregate manager) does not observe all of the actions taken by the agent (the subsystem managers).

Cost analysis: When market prices are known, one can extend the analysis to explore the cost (or profit) efficiency of the various subsystems using models such as those proposed by Färe, Grosskopf, and Lovell (1994). Such an analysis may reveal interesting situations where the market prices of the input and output commodities may differ from the shadow prices obtained from the optimal solution of the models we propose here.

Reallocation: We have shown how it is possible for 2 subsystems to significantly improve their operational efficiencies via reallocation beyond what would be predicted by conventional analysis. However, such improvements may not be realized for certain DMU's, and it may be fruitful to obtain an understanding of when the reallocations proposed herein do not improve operational efficiencies.

Inter-DMUs trading: Our framework allows the manager of a subsystem (say, M_{11}) to trade resources only with the relevant counterpart in the same DMU (here, M_{21}). It might be useful to explore what would happen if resources may be traded among *different* DMUs. (For example, M_{11} acquires resources from M_{24} in return for delivering the amount of intermediate factor (I_4) to which the latter is committed). This might be a prelude to considering mergers or cooperation among independent entities in certain settings.

7. Concluding remarks

DEA is a methodology aimed at evaluating the relative efficiency of DMUs. The construction of composite units that serve as benchmarks against which the performance of observed units are compared lies at the core of DEA. Each DEA model is characterized by a set of assumptions that are translated into a specific mathematical formulation that defines the possible configurations for the composite units, namely, the technology.

We have presented an approach to simultaneously measure the efficiency of aggregate DMUs with two subsystems in series, which goes beyond simply applying standard DEA analysis to each subsystem separately. The main novelty in our proposed approach lies in the more flexible manner in which we model the technology sets for each subsystem by allowing each subsystem to acquire resources from the other and to construct composites from both subsystems.

Our approach is potentially useful in various settings, both at the manufacturing and the service industries. In manufacturing, it can be implemented in the petro-chemical or refinery industries which are characterized by sequential processes where the output of a given stage (e.g., oil refined to a certain Octan level) enters the next stage as an input. Similar processes can be found in food manufacturing plants and other sequential industries. Another area of implementation that bridges the manufacturing and service industries is the warehouse and distribution industry. In many cases, a warehouse may be modelled as a two-stage system in which the first stage uses labor, equipment and space to unload and store inventory, while the second stage uses labor, equipment, space and the inventory to prepare the packages shipped to end-users (see Hackman et al., 2001). An example of a possible implementation in the service industry was pointed out by Chen and Zhu (2004) in the banking context where the first stage uses labor, capital and space to generate deposits and the second stage uses labor, capital and deposits to generate loans.

Recently, there were some attempts to link DEA with current issues in the field of Supply Chain Management (SCM) (see Zhu, 2002). We believe that our approach may further contribute in this context. For example, the implicit competition between S_{1j} and S_{2j} on the weight of the intermediate factor I_j that we describe in Section 5 resembles the *double marginalization* effect that is described in Chapter 5 of Tayur, Ganeshan, and Magazine (1999) where the competition revolves around the pricing of intermediate products. When the chain is owned and operated by a single party (i.e., when vertical integration has been affected) the proposed models could become useful in evaluating the performance of nodes in the chain and making appropriate managerial decisions. For example, the strongest could be allocated additional resources while the activities performed by the weakest nodes could be outsourced.

Acknowledgments The authors wish to express their gratitude to the editor and two anonymous reviewers for their helpful comments.

References

Banker, R.D., A. Charnes, and W.W. Cooper. (1984). "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis." *Management Science*, 30, 1078–1092.

Charnes, A., W.W. Cooper, B. Golany, R. Halek, G. Klopp, E. Schmitz, and D. Thomas. (1986). "Two Phase Data Envelopment Analysis Approaches to Policy Evaluation and Management of Army Recruiting Activities: Tradeoffs Between Joint Services and Army Advertising." Research Report CCS no. 532, Center for Cybernetic Studies, The University of Texas at Austin.

Charnes, A., W.W. Cooper, A.R. Lewin, and L.M. Seiford. (1994). *Data Envelopment Analysis: Theory, Methodology and Applications*. Kluwer Academic Publishers.

Charnes, A., W.W. Cooper, and E. Rhodes. (1978). "Measuring the Efficiency of Decision Making Units." *European Journal of Operational Research*, 2, 429–444.

Charnes, A., W.W. Cooper, and R. Thrall. (1986). "Identifying and Classifying Scale and Technical Efficiencies in Data Envelopment Analysis." *Operations Research Letters*, 5, 105–116.

Chen, Y. and J. Zhu. (2004). "Measuring Information Technology's Indirect Impact on Firm Performance." *Information Technology & Management Journal*, 5, 9–22.

Cook, W., D. Chai, J. Doyle, and R. Green. (1998). "Hierarchies and Groups in DEA." *Journal of Productivity Analysis*, 10, 177–198.

Färe, R., S. Grosskopf, and C.A.K Lovell. (1994). *Production Frontiers*. Cambridge University Press.

Färe, R. and S. Grosskopf. (1996). *Intertemporal Production Frontiers: With Dynamic DEA*. Kluwer Academic Publishers.

Färe, R. and S. Grosskopf. (2000). "Network DEA." *Socio-Economic Planning Sciences*, 34, 35–49.
Färe, R., S. Grosskopf, and S.K. Li. (1992). "Linear Programming Models for Firm and Industry Performance." *The Scandinavian Journal of Economics*, 94, 599–608.
Färe, R. and D. Primont. (1984). "Efficiency Measures for Multiplant Firms." *Operations Research Letters*, 3, 257–260.
Färe, R. and G. Whittaker. (1995). "An Intermediate Input Model of Dairy Production Using Complex Survey Data." *Journal of Agricultural Economics*, 46, 201–213.
Graves, S.C., A.H.G. Rinnoy Kan and P.H. Zipkin. (1993). *Logistics of Production and Inventory*. North-Holland, Amsterdam.
Hackman, S.T., E.H Frazelle, P.M. Griffin, S.O. Griffin, and D.A. Vlatsa. (2001). "Benchmarking Warehousing and Distribution Operations: An Input-Output Approach." *Journal of Productivity Analysis*, 16, 79–100.
Hackman, S.T. and R.C. Leachman. (1989). "A General Framework for Modeling Production." *Management Science*, 35, 478–495.
Hackman, S.T. and U. Passy. (2002). "Maximizing a Linear Fractional Function on a Pareto Efficient Frontier." *Journal of Optimization Theory and Applications*, 113, 83–103.
Hoopes, B., K.P. Triantis, and N. Partangel. (2000). "The Relationship Between Process and Manufaturing Plant Performance: A Goal Programming Approach." *International Journal on Operations and Quantitative Management*, 6, 287–310.
Jehle, G.A. and P.J. Reny. (2001). *Advanced Microeconomic Theory*, 2nd edn. Addison-Wesley, NY.
Johnson, L.A. and D.C. Montgomery. (1974). *Operations Research in Production Planning, Scheduling and Inventory Control*. John Wiley, NY.
Löthgren, M. and M. Tambour. (1999). "Productivity and Customer Satisfaction in Swedish Pharmacies: A DEA Network Model." *European Journal of Operational Research*, 115, 449–458.
Russell, R. (1985). "Measures for Technical Efficiency." *Journal of Economic Theory*, 35, 109–126.
Shephard, R.W. (1970). *Theory of Cost and Production Functions*. Princeton University Press.
Tayur, S., R. Ganeshan, and M. Magazine. (1999). *Quantitative Models for Supply Chain Management*. Kluwer.
Troutt, M.D., P. Ambrose, and C.K. Chan. (2001). "Optimal Throughput for Multistage Input–Output Processes." *International Journal of Operations and Production Management*, 21, 148–158.
Zhu, J. (2002). *Quantitative Models for Performance Evaluation and Benchmarking: Data Envelopment Analysis with Spreadsheets*. Kluwer Academic Publishers, Boston.