Input aggregation bias in technical efficiency with multiple criteria analysis

Casasnovas, Valero L. and Aldanondo, Ana M.

Public University of Navarre

12 June 2014

Input aggregation bias in technical efficiency with multiple criteria analysis

Ana M. Aldanondo [a] & Valero L. Casasnovas [a,*]

[a] Department of Business Administration, Public University of Navarre, Campus Arrosadia, Pamplona, 31006, Spain

* Corresponding author. E-mail: valero.casasnovas@unavarra.es

WORKING PAPER

June 2014

Abstract

We extend the Tauer (2001) and Färe *et al*. (2004) analyses of aggregation bias in technical efficiency measurement to multiple criteria decision analysis. We show input aggregation conditions consistent with multiple criteria evaluation of overall efficiency in conjunction with variation in aggregation bias.

Keywords: Data Envelopment Analysis, Input aggregation, multiple objectives.

JEL Codes: C61, D20

I. Introduction

Aggregate input values are a common feature in Data Envelopment Analysis (DEA) applications. Researchers estimating the technical efficiency of different economic units often use raw material costs rather than quantities per material and labour costs instead of the number of hours worked per job category. While there are advantages in using aggregate data to overcome information gaps or reduce the dimensions of the problem, there are also some drawbacks. Various studies (Primont, 1993; Thomas and Tauer, 1994; Tauer, 2001; Färe and Zelenyuk, 2002; Färe *et al.*, 2004) suggest that measures of radial technical efficiency using linear input aggregates could be biased by price allocation inefficiency.

The main approach to the theoretical analysis of technical efficiency aggregation bias is based on the standard model of cost efficiency and its decomposition into technical efficiency and allocation efficiency. In this context, cost minimization (as a prerequisite for profit maximization) is the only economic aim of the production unit. The question is whether this scenario is always warranted. Indeed, in complex production problems, decisions require a combination of multiple, conflicting optimization criteria and efficiency. For instance, a plant manager may seek simultaneously to reduce production costs, optimize management time and reduce pollutant inputs (Coelli *et al.*, 2007). Thus, we have multiple overall efficiency criteria and a single technology (or technical efficiency). This may pose an additional problem of inconsistency when estimating technical efficiency using aggregate inputs. Technical efficiency estimates based on a single composite input (or materials,…) price may be incompatible with overall environmental efficiency (or cost,…) estimates and other criteria.

The main purpose of this paper is to extend the analysis of input aggregation bias[1] to situations in which the production units are simultaneously focused on several objectives. The two main issues we wish to address are a) the number and nature of the aggregate values required in order to avoid inconsistent estimates; and b) how far the degree of bias in the efficiency score varies when using several linear aggregates of the same inputs.

II. Presentation of the different models

Let us consider a situation in which we observe the production activity of $k=1,\ldots,K$ firms using N inputs, $x^k= (x_{k1},x_{k2,\ldots,}x_{kN}) \in R_N^+$ in order to produce M outputs $y^k= (y_{k1}, y_{k2,\ldots,}y_{kM}) \in R_M^+$ using the same linear technology. We assume that the technology is regular and presents variable returns to scale (VRS).[2] The production possibility set is considered defined by the convex combination of these observed activities (Färe *et al.*, 1994). Here, let us recall the definition of the production input set $L(y)$ as the set of inputs vectors, $x \in R_N^+$ yielding at least output level $y \in R_M^+$ (Shepard, 1970; Färe *et al.*,1994).

Inputs have their corresponding prices $(w_{1,}w_{2,\ldots,}w_n) \in R_N^+$, which are assumed to be the same for all firms. To illustrate the problem,[3] let us further suppose that the production process also generates $h=1,\ldots,H$ pollutant residues. Each unit of input, n, includes a quantity of different materials $(a_{n1,}a_{n2,\ldots,}a_{nH}) \in R_H^+$ some of which are discharged into nature as pollutant residues.[4] The total material h used by firm k is given by the following equation:

[1] Similar analysis is possible for aggregate outputs.
[2] The analysis can be extended to constant returns to scale.
[3] We use the environmental condition to give meaning to the second set of input coefficients. This can be generalised to other problems
[4] For more on materials balance-based models of environmental efficiency, see Coelli *et al*. (2007). See also Hampf (2014) for a justification of the strong disposability condition in material balance models.

$$A_{kN}^{h} = \sum_{n=1}^{N} a_{nh} x_{kn} \qquad h = 1,\ldots,H \qquad (1)$$

Where a_{nh} is the unit coefficient of substance h for input n, which may be null for some inputs. For the sake of simplicity, we will begin by assuming that the productive activity generates a single pollutant residue, $H=1$ allowing us to drop the index h in the notation. Therefore, an increase in technical efficiency in inputs also implies a decrease in costs and a gain in overall environmental efficiency. We will now consider how the radial technical efficiency score (Farrell, 1957) varies with an aggregation of $\hat{N} \leq N$ inputs, simultaneously using two types of linear aggregates of the same inputs, one price-weighted, w_n, and another weighted by the material flow coefficients, a_n.

$$C_{k\hat{N}} = \sum_{n=1}^{\hat{N}} w_n x_{kn} \qquad k=1,2,\ldots,K \qquad \hat{N} \leq N \qquad (2)$$

and

$$A_{k\hat{N}} = \sum_{n=1}^{\hat{N}} a_n x_{kn} \qquad k = 1,2,\ldots,K \qquad \hat{N} \leq N \qquad (3)$$

Starting from previous results obtained by Färe *et al.* (2004), we will compare the input-oriented technical efficiency scores estimated for firm k using a single linear aggregate of inputs (costs or materials, indistinctly) with that estimated using two aggregates simultaneously. To this end, we will define four efficiency scores for the same production unit k using the data envelopment analysis (DEA) model of Banker, Charnes and Cooper (1984), BCC, for technologies with variable returns to scale.

$EFi(y^k,x^k)$, which is the estimated efficiency score, is calculated without aggregate input values, by the following linear program:

$$
\begin{aligned}
& EFi\left(y^k, x^k\right) = \min\ \beta \\
& \text{subject} \quad \text{to} \\
& \sum_k z_k y_{km} \geq y_{km} \qquad m = 1,\ldots, M \\
& \sum_k z_k x_{kn} \leq \beta x_{kn} \qquad n = 1,\ldots, N \\
& z_k \geq 0, \quad \sum_k z_k = 1 \qquad k = 1,\ldots, K
\end{aligned} \tag{4}
$$

The efficiency of unit k using aggregate costs of $\hat{N}$ inputs, $EF_C(y^k, C_{k\hat{N}}, x_{k\hat{N}+1},\ldots, x_{kN})$ is estimated by the following linear program:

$$
\begin{aligned}
& EF_C\left(y^k, C_{k\hat{N}}, x_{k\hat{N}+1},\ldots, x_N\right) = \min\ \beta \\
& \text{subject} \quad \text{to} \\
& \sum_k z_k y_{km} \geq y_{km} \qquad m = 1,\ldots, M \\
& \sum_k z_k C_{k\hat{N}} \leq \beta C_{k\hat{N}} \\
& \sum_k z_k x_{kn} \leq \beta x_{kn} \qquad n = \hat{N}+1,\ldots, N \\
& z_k \geq 0, \quad \sum_k z_k = 1 \qquad k = 1,\ldots, K
\end{aligned} \tag{5}
$$

We calculate the efficiency score using aggregate material inputs $EF_A\left(y^k, A_{k\hat{N}}, x_{k\hat{N}+1},\ldots, x_{kN}\right)$ by the same program as in (5), substituting $C_{k\hat{N}}$ with $A_{k\hat{N}}$.

We then extend program (5) to include both $C_{k\hat{N}}$ and $A_{k\hat{N}}$ as inputs, in order to obtain the efficiency score using double aggregation:

$$EF_{CA}\left(y^k, C_{k\hat{N}}, A_{k\hat{N}}, x_{k\hat{N}+1},..., x_N\right) = \min\ \beta$$

subject to

$$\begin{aligned}
&\sum_k z_k y_{km} \geq y_{km} && m = 1,..., M \\
&\sum_k z_k C_{k\hat{N}} \leq \beta C_{k\hat{N}} && \\
&\sum_k z_k A_{k\hat{N}} \leq \beta A_{k\hat{N}} && \\
&\sum_k z_k x_{kn} \leq \beta x_{kn} && n = \hat{N}+1,..., N \\
&z_k \geq 0, \quad \sum_k z_k = 1 && k = 1,..., K
\end{aligned} \tag{6}$$

Thus, we have four DEA measures of input-oriented technical efficiency:[5] *EFi,EFc,* EF_A, EF_{CA}. The second and third use a single linear aggregate as the input (the subcost or sub-pollutant), the fourth uses two aggregates of the same inputs.

III. Technical efficiency biases using multiple aggregates

We will begin by showing that the technical efficiency score is consistent and less biased when the estimation uses two linear aggregates of the same inputs (costs and environmental load) than when it uses a single aggregate. For this, we start from previous results obtained by Färe *et al*. (2004) for an analysis using a single linear aggregate. These authors show that, for every observation *k (k=1,…,K)*, it is satisfied that:

$$\begin{gathered}
EF_C\left(y^k, C_{kN}\right) \leq EF_C\left(y^k, C_{k\hat{N}}, x_{k\hat{N}+1},...,x_N\right) \leq EFi\left(y^k, x^k\right) \\
\text{and} \\
EF_A\left(y^k, A_{kN}\right) \leq EF_A\left(y^k, A_{k\hat{N}}, x_{k\hat{N}+1},...,x_N\right) \leq EFi\left(y^k, x^k\right)
\end{gathered} \tag{7}$$

Where C_{kN} and A_{kN} are, respectively, the observed firm´s cost and material use; EF_C (y^k,C_{kN}) represents the variable returns to scale (VRS) cost efficiency (Färe and

[5] Similar analysis is possible for aggregate outputs.

Grosskopf, 1985); and EF_A (y^k,A_{kN}) is the VRS overall environmental efficiency. Then, when some inputs are linearly aggregated while others are kept disaggregated, the technical efficiency score given by the BCC model exhibits a downwards bias with respect to that given by an application using fully unbundled inputs, and this bias is limited by the allocative efficiency (Färe *et al.*, 2004).

According to these results, a unit that performs well in terms of overall environmental efficiency can be classed as technically inefficient if the calculations are made using program (5) with a single aggregate cost variable. For this to occur, the unit needs to be cost allocation inefficient. Similar mis-measurement can occur with a unit that is overall cost-efficient if the estimate is computed using only aggregate material inputs.

Applying the same conditions as in equations (7), as will be seen later, it can be proven that, for any k ($k=1,\ldots,K$), it is satisfied that:

$$EF_C\left(y^k, C_{k\hat{N}}, x_{k\hat{N}+1},\ldots,x_N\right) \leq EF_{CA}\left(y^k, C_{k\hat{N}}, A_{k\hat{N}}, x_{k\hat{N}+1},\ldots,x_N\right) \leq EFi\left(y^k, x^k\right)$$

and (8)

$$EF_A\left(y^k, A_{k\hat{N}}, x_{k\hat{N}+1},\ldots,x_N\right) \leq EF_{CA}\left(y^k, C_{k\hat{N}}, A_{k\hat{N}}, x_{k\hat{N}+1},\ldots,x_N\right) \leq EFi\left(y^k, x^k\right)$$

In other words, the technical efficiency score estimated by linear program (6), using two linear aggregates of the same inputs, one for costs (equation 2) the other for materials (equation 3) and non-aggregated values of all other variables, is biased downwards in relation to the estimate given by program (4).The degree of bias is less than or equal to that which occurs in the measure of technical efficiency using a single linear aggregate, whether of costs or materials. Therefore the use of both aggregates together cannot possibly give the inconsistent result of finding a unit overall cost- or environment-efficient and at the same time technically inefficient.

To prove the relationship in equations (8), one only needs to demonstrate, in a way analogous to that used by Färe *et al.* (2004), that the set of all feasible solutions given by linear program (4) is a subset of the set of feasible solutions given by linear program (6).[6] Therefore, the smallest optimal solution given by program (4) must be greater than or equal to that given by program (6). Linear program (6) is, in turn, a more constrained version of program (5). Thus, the smallest optimal solution given by program (6) will be no smaller than that given by program (5), which proves the relationship in equations (8).This result can be generalised to the case using more than two linear aggregates of the same inputs, $H>1$. An increase in the number of linear aggregates of the same inputs used in the DEA technical efficiency estimation reduces potential aggregation bias.

Next, we will prove that a sufficient condition for zero aggregation bias in technical efficiency estimated by linear program (6) is that the input of the k^{th} unit should be a positive linear combination of the inputs of an overall variable return to scale (VRS) cost-efficient unit, $c(y^c,x^c)$ (Färe and Grosskopf, 1985) and an overall VRS environmentally-efficient unit, $a(y^a,x^a)$ as indicated by the following expression:[7]

$$\begin{aligned} & y_{km} = y_{cm} = y_{cm} && m = 1,..., M \\ & x_{kn} = \alpha_c x_{cn} + \alpha_a x_{an} && n = 1,..., N \\ & \alpha_c \geq 0, \quad \alpha_a \geq 0, \quad \alpha_c + \alpha_a \geq 1 && \\ & x^k \in L(y^k) && \end{aligned} \tag{9}$$

Where index c denotes the overall cost-efficient unit and index a denotes the overall environmentally-efficient unit (Coelli *et al.*, 2007).

Let us first consider the benchmark case, in which $\hat{N} = N$. That is, where technical efficiency is estimated by program (6) with only two inputs: the firm's observed cost and observed material inputs, $EF_{CA}\ (y^k, C_{kN,}, A_{kN})$. The purpose of our proof is to demonstrate that the optimal solution to linear programs (4) and (6) is found with no

---

[6] The proof, omitted for reasons of space and because the procedure is identical to that of Primont (1996) and Färe *et al.* (2004), is available upon request from the authors. Program (3) has arguably more constraints than program (4) because the input constraint has to be satisfied for all prices.

[7] For a determinate y^k or scale of production because we assume variable returns to scale (Tone, 1996; Krivonozhko and Førsund, 2010)

slacks in input restrictions for the k^{th} unit; that is, the k^{th} unit is enveloped with a subset of efficient units of the input set $L(y^k)$ (Shepard, 1970; Färe *et al.*, 1983). Here, it can be recalled that BCC-efficient units (Tone, 1996; Cooper *et al.*, 2007) are those for which the optimal solution given by the BCC program is equal to 1 with no slacks in input and output restrictions. Now, we can define as BCC input efficient units those with an efficiency score equal to unity and no slacks in program (4) input restrictions,[8] that is, no unit in the input set for the output of the BCC input-efficient unit has a smaller input vector than the BCC input efficient unit (Shepard, 1970; Färe *et al.*, 1983).

We begin with linear program (4). Let it be noted that units, c and a are both BCC-input-efficient for program (4), because they are run with minimum input costs and minimum material inputs, respectively, for output y^k (Sueyoshi, 1999).[9] Furthermore, the convexity of technology determines that input set $L(y^k)$ includes a BCC-input-efficient unit which has an input vector that is a positive linear combination of the inputs of the two efficient units, c and a, as indicated by the following expression:[10]

$$\lambda_x\left(\frac{\alpha_c}{\alpha_c+\alpha_a}x^c+\frac{\alpha_a}{\alpha_c+\alpha_a}x^a\right)\in L\left(y^k\right) \tag{10}$$

where $\lambda_x \leq 1$.

Therefore, for the k^{th} unit, projected efficiency according to program (4) (z^*, y^*, x^*, β^*) is written as follows:

$$\begin{aligned}
&x_n^* = \sum_k z_k^* x_{kn} = \beta^*\left(\alpha_c x_{cn} + \alpha_a x_{an}\right) && n = 1,\ldots,N \\
&y_{km}^* = \sum_k z_k^* y_{km} \geq y_{km} && m = 1,\ldots,M \\
&z_k^* \geq 0, \quad \sum_k z_k^* = 1 && k = 1,\ldots,K
\end{aligned} \tag{11}$$

[8] Tone and Tsutsui (2010) propose an equivalent category in a constant returns to scale model.
[9] This condition can be proved by comparing the solutions of the dual of programs (4) and (5). The proof is omitted here but is available upon request from the authors.
[10] Actually, all efficient units that are positive combinations of a and c are overall efficient for both criteria cost and environment together. And inefficient units enveloped by this part of the frontier are allocation- efficient.

Units c and a, for their part, are BCC input-efficient (Kornbluth, 1974) by linear program (6); because they have the output y^k minimal production cost and lower environmental impact, respectively. Therefore, as for linear program (4), the projected activity of linear program (6) (z^{**}, y^{**}, C^{**}, A^{**}, β^{**}) is written as follows:

$$\begin{aligned}
C_{kN}^{**} &= \sum_k z_k^{**} C_{kN} = \beta^{**}\left(\alpha_c C_{cN} + \alpha_a C_{aN}\right) \\
A_{kN}^{**} &= \sum_k z_k^{**} A_{kN} = \beta^{**}\left(\alpha_c A_{cN} + \alpha_a A_{aN}\right) \\
y_{km}^{**} &= \sum_k z_k^{**} y_{km} \geq y_{km} \qquad m = 1,\ldots, M \\
z_k^{**} &\geq 0, \quad \sum_k z_k^{**} = 1 \qquad k = 1,\ldots, K
\end{aligned} \tag{12}$$

Given that (z^*,x^*) is that unit of the input set which maximally proportionally contracts the inputs of the k^{th} unit, it is satisfied that $\beta^*=\beta^{**}$.[11]

Finally, it can be shown, in a way analogous to that used by Färe *et al*. (2004), that the following relationship is satisfied:

$$EF_{CA}(y^k, C_{kN}, A_{kN}) \leq EF_{CA}(y^k, C_{k\hat{N}}, A_{k\hat{N}}, x_{k\hat{N}+1}, \ldots, x_N) \tag{13}$$

It only has to be proved that the set of feasible solutions to program (6) for the case in which $\hat{N} < N$ is a subset of the set of feasible solutions to program (6) when $\hat{N} = N$.

Then, if equations (11), (12) and (13) are satisfied, it is proved that there is zero aggregate bias in program (6) when the unit under analysis is a positive linear combination of a cost-efficient unit and an environmentally-efficient unit.

These results can be generalised to the case in which more than two linear aggregates of the same inputs are used.

The overall conclusion of this study is that aggregate bias in technical efficiency estimation decreases with the used number of linear aggregates of the same inputs and is null for allocation-efficient units (in terms of prices, environment,…) and for units

[11] It is sufficient to write *(C_{cN}, A_{cN})* and *(C_{aN}, A_{aN})* as a function of the inputs of units a and c in (11). Further proof is available upon request from the authors.

that are a positive combination of a number of overall efficient units. These results have important implications for applied analysis. First, DEA technical efficiency programs with composite inputs should use as many linear aggregates of inputs as overall efficiency criteria to avoid inconsistencies between technical efficiency and overall efficiency estimates. Second, an adequate use of multiple linear aggregates could significantly reduce technical efficiency aggregation bias.

References

Banker, R. D., Charnes, R. F. and Cooper, W. W. (1984) Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis, *Management Science*, 30, 1078–92.

Coelli, T., Lauwers, L. and Van Huylenbroeck, G. (2007) Environmental efficiency measurement and the materials balance condition, *Journal of Productivity Analysis*, 28, 3-12.

Cooper, W. W., Seiford, L. M. and Tone, K. (2007) *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*, Second Edition, Kluwer Academic Press, Boston.

Färe, R. and Grosskopf, S. (1985) A Nonparametric Cost Approach to Scale Efficiency *The Scandinavian Journal of Economics*, 87, 594-604.

Färe, R., Grosskopf, S. and Lovell, C. K. (1983) The Structure of Technical Efficiency, *The Scandinavian Journal of Economics*, 85, 181-90.

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

Färe, R., Grosskopf, S. and Zelenyuk, V. (2004) Aggregation bias and its bounds in measuring technical efficiency, *Applied Economics Letters*, 11, 657-60.

Färe, R. and Zelenyuk, V. (2002) Input aggregation and technical efficiency, *Applied Economics Letters*, 9, 635-36.

Farrell, M. J. (1957) The measurement of productive efficiency, *Journal of the Royal Statistical Society*, Series A (General), 120, 253-90.

Hampf, B. (2014) Separating environmental efficiency into production and abatement efficiency: a nonparametric model with application to US power plants, *Journal of Productivity Analysis*, 41, 457-73.

Krivonozhko, V. E. and Førsund, F. R. (2010) The relationship between returns-to-scale properties of interior points and vertex points in DEA models, *Doklady Mathematics*, 81, 159-63.

Kornbluth, J. S. H. (1974) Duality, Indifference and Sensitivity Analysis in Multiple Objective Linear Programming, *Operational Research Quarterly*, 25, 559-614.

Primont, D. (1993) Efficiency measures and input aggregation, in *Mathematical Modelling in Economics. Essays in Honor of Wolfgang Eichhorn*, (Eds.) W. E. Diewert, K. Spremann and F. Stehling, Springer, Berlin, pp.288–94.

Shephard, R. W. (1970) *Theory of cost and production functions*, Princeton University Press, Princeton, NJ.

Sueyoshi, T. (1999) DEA Duality on Returns to Scale (RTS) in Production and Cost Analyses: An Occurrence of Multiple Solutions and Differences between Production-Based and Cost-Based RTS Estimates, *Management Science,* 45, 1593-608

Tauer, L. W. (2001) Input aggregation and computed technical efficiency, *Applied Economics Letters*, 8, 295-7.

Thomas, A. C. and Tauer, L. W. (1994) Linear input aggregation bias in nonparametric technical efficiency measurement, *Canadian Journal of Agricultural Economics*, 42, 77-86.

Tone, K. (1996) A Simple Characterization of Returns to Scale in Data Envelopment Analysis. *Journal of the Operations Research Society of Japan*, 39 (4), 604-13.

Tone, K. and Tsutsui, M. (2010) An epsilon-based measure of efficiency in DEA – A third pole of technical efficiency, *European Journal of Operational Research,* 207, 1554-63.