Aggregation of Scale Elasticities Across Firms

Fare, Rolf and Zelenyuk, Valentin

Universite Catholique de Louvain, Oregon State University, EERC, UPEG

15 April 2004

AGGREGATION OF SCALE ELASTICITIES ACROSS FIRMS

Rolf Färe[*] and Valentin Zelenyuk[**]

Abstract

In this paper we propose new aggregate or 'group' primal and dual *scale elasticity* measures of an economic system (e.g., industry consisting of several firms, etc.). The main contribution of the paper is that we show under what assumptions a formal relationship between these new aggregate scale elasticity measures and the individual scale elasticities can be established. The derivations are based on the principles of economic optimization, and in particular on the duality theory of Shephard (1953) and the aggregation theory of Koopmans (1957).

Key Words: Scale Elasticity, Aggregation, Duality

JEL Code: D24

April 2010

[*] — Department of Economics and Department of Agricultural and Resource Economics, Oregon State University (Corvallis, OR 97331, USA.), USA. e-mail: rolf.fare@orst.edu

[**] — (✉) School of Economics and Centre for Efficiency and Productivity Analysis, The University of Queensland, Australia, 530, Colin Clark Building (39), St Lucia, Brisbane, Qld, 4072, AUSTRALIA; v.zelenyuk@uq.edu.au; tel: + 61 7 3346 7054.

Acknowledgement: We would like to thank Tom Coupé, Roy Gardner, Volodymyr Vakhitov and anonymous referees for their comments. Research support from the "Interuniversity Attraction Pole", Phase V (No.P5/24) from the Belgian Government (Belgian Science Policy) is acknowledged. All views expressed remains responsibility of the authors.

1. Introduction

Applied economists are often interested in measuring economies of scale present in an economic system (e.g., industry consisting of several firms, etc.). To do so, researchers often resort to estimating the scale elasticity *at* the mean of the data or, alternatively, take the equally-weighted mean of the *individual* (for each observation) estimates of scale elasticities. Both measures have some theoretical appeal, although not equal in general. In this study we propose another theoretical measure of the scale elasticity of a *group* of independent decision making units (DMUs)—the analogue to *individual* scale elasticity measure but defined on the aggregate technology. We then show under which assumptions a formal relationship between these new aggregate scale elasticity measures and the individual scale elasticities can be established. The derivations are based on the principles of economic optimization, and in particular on the duality theory of Shephard (1953) and the aggregation theory of Koopmans (1957).

The rest of the paper is structured as follows. In Section 2, we outline the main definitions for the technology characterizations and scale elasticity measurement of an individual firm. In Section 3, we introduce the notions of aggregate or group technology characterization and the related aggregate or group scale elasticity measurement and then derive the relationship to their individual analogues. In Section 4 we discuss the issue of price independent weights. In Section 5 we discuss limitations of the presented results and noting on possible extensions and future work. Section 6 concludes.

2. Individual Technology Characterization and Scale Elasticity Measures

Consider an example of an economic system (e.g., industry) where each decision making unit (DMU, e.g., a firm) $k \in \{1, \ldots, K\}$ uses vector of N inputs, denoted with $x^k = (x_1^k, \ldots, x_N^k)^T \in \Re_+^N$, to produce vector of M *outputs*, denoted by $y^k = (y_1^k, \ldots, y_M^k)^T \in \Re_+^M$. Following Shephard (1953) and Färe and Primont (1995), suppose technology of a firm $k \in \{1, \ldots, K\}$ can be characterized by the input requirement set,

$$L^k(y^k) \equiv \{x: \ x \ can \ \ produce \ \ y^k\}, \qquad y^k \in \Re^M_+ \tag{2.1}$$

which we assume satisfies the standard regularity axioms of production theory (see Färe and Primont (1995) for details), so that this technology can be characterized (dually) via the *cost* function,

$$C^k(y^k, \ w) \equiv \min_x\{wx \quad : \quad x \in L^k(y^k)\}, \tag{2.2}$$

where $w \equiv (w_1,...,w_N) \in \Re^N_{++}$ denotes the vector of input prices. In a single-output case (*M=1*) technology of a firm *k* can be characterized by the *production function*. For the multi-output case (*M>1*), we can use the *input orientated* Shephard's (1953) distance function $D_i^k : \Re^M_+ \times \Re^N_+ \to \Re^1_+ \cup \{\infty\}$,

$$D_i^k(y^k, x^k) \equiv \sup_\theta\{\theta > 0: \ (x^k / \theta) \in L^k(y^k)\}, \tag{2.3}$$

which gives an alternative (primal) characterization of technology of the firm *k*, in the sense that

$$D_i^k(y^k, x^k) \geq 1 \quad \Leftrightarrow \quad x^k \in L^k(y^k). \tag{2.4}$$

Both the *dual*, given in (2.2), and the *primal*, given in (2.3), characterizations are often used in measuring properties of a technology. The focus of our paper is on the *scale economies*, which can be inferred from the notion of *scale elasticity*, defined for the *dual* and *primal* frameworks as (Färe and Primont, 1995):

$$e_c(y^k, w) \equiv \left. \frac{\partial \ln C^k(y^k\theta, w)}{\partial \ln \theta} \right|_{\theta=1} = \frac{\nabla^T_{y^k} C^k(y^k, w) y^k}{C^k(y^k, w)} \quad , \tag{2.5}$$

and

$$e_i(y^k, x^k) \equiv \left. \frac{\partial \ln \lambda}{\partial \ln \theta} \right|_{\theta=1, \lambda=1} = -\nabla^T_{y^k} D_i^k(y^k, x^k) y^k \ \text{ such that } \ D_i^k(y^k\theta, x^k\lambda) = 1 \tag{2.6}$$

respectively. Letting x^{*k} be a solution to (2.2), we get the dual and primal measures being equal, i.e.,

$$e_c(y^k, w) = e_i(y^k, x^{*k}), \qquad x^{*k} \equiv \arg\min_x\{wx \quad : \quad x \in L^k(y^k)\}. \tag{2.7}$$

This result states that the same information about the scale elasticity of an *individual* firm k can be obtained from primal and dual approaches. In the next section we show analogous results for the measures of scale elasticity of a *group* and its relationship to these individual measures.

3. Aggregation of Technology and of Scale Elasticity

Define the *group* (e.g., industry, sub-industry, etc) input requirement set to be

$$\overline{L}(y^1,..., \ y^K) = \sum_{k=1}^{K} L^k(y^k). \tag{3.1}$$

The *group* cost function, which is the *group* analog of the (2.2), can be defined as

$$\overline{C}(y^1,..., y^K, \ w) \equiv \min_x \{wx \ : \ x \in \overline{L}(y^1,..., y^K)\}, \tag{3.2}$$

and the *group* input oriented distance function—the *group* analog of (2.3)—can be defined as

$$\overline{D}_i(y^1,..., y^K, \sum_{k=1}^{K} x^k) \equiv \sup_\theta\{\theta > 0: \ (\sum_{k=1}^{K} x^k / \theta) \in \overline{L}(y^1,..., y^K)\}. \tag{3.3}$$

The *scale economies* for the *group* can be inferred from measures of scale elasticities defined for the *aggregate* technology in the same manner as for the individual technologies. Thus, for the dual framework we have

$$\overline{E}_c(y^1,..., y^K, w) \equiv \left. \frac{\partial \ln \overline{C}^k(y^1\theta,..., y^K\theta, w)}{\partial \ln \theta} \right|_{\theta=1} = \frac{\nabla_Y^T \overline{C}(y^1,..., y^K, w)Y}{\overline{C}(y^1,..., y^K, w)}, \tag{3.4}$$

where $\nabla_Y^T \overline{C}(y^1,..., y^K, w) \equiv (\partial \overline{C}(y^1,..., y^K, w) / \partial y^1, \ldots, \partial \overline{C}(y^1,..., y^K, w) / \partial y^K)$, and $Y \equiv (y^1,..., y^K)^T$.

On the other hand, for the primal framework, given $\overline{D}_i(y^1\theta,..., y^K\theta, \sum_{k=1}^{K} x^k\lambda) = 1$, we have

$$\overline{E}_i(y^1,..., y^K, \sum_{k=1}^{K} x^k) \equiv \left. \frac{\partial \ln \lambda}{\partial \ln \theta} \right|_{\theta=1,\lambda=1} = -\nabla_Y^T \overline{D}_i(y^1,..., y^K, \sum_{k=1}^{K} x^k)Y. \tag{3.5}$$

Let x^* be a solution to (3.2), then the dual and primal measures of group scale elasticities are equal, i.e.,

$$\overline{E}_c(y^1,..., y^K, w) = \overline{E}_i(y^1,..., y^K, x^*), \tag{3.6}$$

where

$$x^* \equiv \arg\min_x \{wx \quad : \quad x \in \overline{L}\ (y^1,..., y^K)\}.$$

This is an analog to the relationship (2.7) that was derived for the individual level by Färe and Primont (1995) and for the group level it is established in a similar manner.

We now want to establish the relationship between the aggregate (group) and the individual scale elasticities and the key instrument for this would be the following theorem.

Theorem. Given definitions (2.1), (2.2), (3.1) and (3.2), we have

$$\overline{C}(y^1,..., \ y^K, \ w) = \sum_{k=1}^{K} C^k(y^k, w). \tag{3.7}$$

In words, the group (e.g., industry) cost function is the sum of individual cost functions of all DMUs (e.g., firms) in this industry. This theorem is from Färe, Grosskopf and Zelenyuk (2004) and is the cost analog of the Koopmans (1957) theorem for profit functions. (Färe and Zelenyuk (2003) provide revenue analog. Also see Kuosmanen et al. (2006) and Zelenyuk (2006, 2010) for use of this result in other contexts.) It might be worth noting that despite the fact that (3.7) looks quite intuitive it shall not be accepted as a universal truth that one could get without going through derivations. Indeed, the fundamental result (3.7) relies on explicit assumptions: (i) about the structure of aggregate technology (additive in input requirement sets), stated in (3.1), (ii) the law of one price for inputs, and (iii) optimization behavior. If another aggregation structure is assumed (e.g., summation of output sets rather than input sets or, alternatively, summation of technology sets, etc.) then a different aggregation result might emerge. We return to this issue in the next section.

Meanwhile, by treating expression (3.7) as an identity we can differentiate both sides of it along the ray from the origin through the point $Y \equiv (y^1,..., y^K)^T$ (i.e., looking at the change in costs due to infinitesimal *equi-proportional* change of *all* outputs) and obtain the desired aggregation result. Specifically, such differentiation of the l.h.s. of (3.7) along the ray gives

$$\left.\frac{\partial \overline{C}(y^1\theta,\ldots,\ y^K\theta,\ w)}{\partial\theta}\right|_{\theta=1} = \nabla_Y^T \overline{C}(y^1,\ldots,\ y^K,\ w)Y\,. \tag{3.8}$$

And, differentiating the r.h.s. along the same ray gives

$$\left.\partial\left(\sum_{k=1}^{K} C^k(y^k\theta, w)\right)\Big/\partial\theta\right|_{\theta=1} = \sum_{k=1}^{K} \nabla_{y^k}^T C^k(y^k, w)\, y^k \quad . \tag{3.9}$$

Putting the two sides together yields the desired result:

$$\overline{E}_c(y^1,\ldots,\ y^K,\ w) = \sum_{k=1}^{K} e_c(y^k, w)\cdot S^k\,, \qquad S^k \equiv C^k(y^k, w)\Big/\sum_{k=1}^{K} C^k(y^k, w) \tag{3.10}$$

This mathematical fact has quite intuitive economic appeal: The dual scale elasticity of the group can be obtained as the *weighted* sum of the *individual* dual scale elasticities of all firms in this group, where the weights are the cost shares. It is worth emphasizing here again that the weights in (3.10) are not ad hoc, but derived from agents' optimization behavior and other explicit assumptions mentioned above, and change of these assumptions might lead to the change of the aggregation result, as we discuss more about it in the next section.

We can also obtain the aggregation result for the *primal* measures of scale elasticities (based on the group and individual input distance functions). By recalling (2.7) and (3.6), we have

$$\overline{E}_i(y^1,\ldots,\ y^K,\ x^*) = \sum_{k=1}^{K} e_i(y^k, x^{*k})\cdot S^k\,, \qquad S^k \equiv C^k(y^k, w)\Big/\sum_{k=1}^{K} C^k(y^k, w)\,. \tag{3.11}$$

where $x^* \equiv \arg\min_x\{wx \ :\ x \in \overline{L}\ (y^1,\ldots, y^K)\}$, and $x^{*k} \equiv \arg\min_x\{wx \ :\ x \in L^k(y^k)\}$.

The result (3.11) gives us a practical way of obtaining the (primal) *group* scale elasticity measure immediately from the (primal) *individual* scale elasticity measures. Specifically, it states that the primal group or aggregate scale efficiency measure can be obtained by taking the *weighted* arithmetic average of individual scale elasticities of all firms in this group, where the weights are the cost shares—the same as in the dual framework. Again, although simple, the aggregation function and weights are not ad hoc but derived from economic principles, with explicit assumptions on aggregate technology, equality of input prices across firms and cost minimizing behavior.

4. Price Independent Weights

When the primal information is used, a researcher may have no price information available to obtain the cost functions used to compute the weights. A feasible option then might be to use the shadow prices, estimated from the primal information. Alternatively, the unavailability of price information can be circumvented by deriving the price independent weights in the manner similar to Färe and Zelenyuk (2003, 2007). This would require imposing additional standardization,

$$w_i \sum_{k=1}^{K} x_i^{*k} \Big/ \left(\sum_{i=1}^{N} w_i \sum_{k=1}^{K} x_i^{*k} \right) = a_i , \qquad i = 1, \ldots , N, \tag{4.1}$$

where a_i is a known (or estimated) constant between zero and unity. Intuitively, (4.1) says that the share of the *industry* expenditures on input i in the industry total cost equals a_i. Now, letting $\varpi_i^k = x_i^k / \sum_{k=1}^{K} x_i^k$ be the share of *k's* firm in the group in terms of i^{th}-input, and using (3.11) with the standardization (4.1) we obtain the price-independent weights, with an intuitive economic meaning,

$$S^k = \sum_{i=1}^{N} \varpi_i^k a_i , \qquad k = 1, \ldots , K. \tag{4.2}$$

Intuitively, (4.2) says that firm's weight is the weighted average over all input-shares of this firm in the group, where the weights are the shares of the industry expenditures on input i in the industry total cost.

5. Limitations, Alternatives and Possible Extensions

As was mentioned above, researchers sometimes summarize their results by presenting the simple (equally-weighted) averages of scale elasticities. This statistic is certainly useful as an estimator of the population mean of scale elasticity distribution for a given population. Yet, because the scale elasticities are normalized quantities, and as a result they ignore the size of the unit that generated it, the equally-weighted averages might give distorted picture of the reality. For this reason the aggregation scheme with weights accounting for economic importance of each unit entering the aggregation derived in this paper must be very useful.

The aggregation function and the weights that came out from the derivations in this paper turned out to be quite intuitive. Yet, the value of the paper is not in proposing a new aggregation scheme per se, but rather in justifying and explaining a one with precise assumptions and formal derivations that have economic meaning (e.g., optimization, etc.). The resulting aggregation scheme, however simple or complicated it is, would then be *not* ad hoc, but based on some theoretical grounds, with clear implications and restrictions on its use. Remarkably, note that the aggregation solution derived here not only does not impose a particular functional form for the production or distance or cost (revenue) function but also allows for totally different technologies underlying those functions. However, there are three important assumptions underlying the resulting aggregation scheme (3.10)-(3.11) and let us recall and discuss each assumption in more detail.

The first assumption states that the aggregate technology structure must satisfy (3.1). This assumption must not be confused with the union of the input requirement sets, but rather understood as the total requirements for inputs to produce certain levels of outputs by each individual decision making unit (DMU). That is, a reallocation of outputs to be produced, given by $(y^1,..., y^K)$, across the DMUs is not allowed, while reallocation of inputs is allowed. As a result, the use of the presented aggregation scheme would not be adequate for the case when there is a 'central planner' of the group of DMUs that can decide on reallocation of outputs to be produced across the DMUs. One example of such case is an industry in a socialist-type economy, where the central planner can reallocate the output plans across firms in the industry. Another example could be when the economic system under investigation is a firm consisting of several plants and the central management of the firm has power to reallocate the output plans across the plants. For these cases, different assumption on aggregate technology might be more appropriate.[1]

One approach to allow for the reallocation of outputs would be to use the elasticities based on revenue functions and Shephard's output distance functions, then one can assume aggregation technology as the sum of output sets (as in Färe and Zelenyuk (2003)), and so allow for cross-firm reallocation of outputs (although not of inputs). The derivations would be analogous to those presented above.

Another approach to allow for the reallocation of outputs is to assume that aggregate technology is the sum of the technology sets (as in Koopmans (1957), Li and Ng (1995) and Nesterenko and Zelenyuk (2007)) and thus allow for cross-firm reallocation of both inputs and outputs. In this case, a different aggregation scheme might emerge (e.g., for the context of aggregation of technical efficiency, Nesterenko and Zelenyuk (2007) arrived to harmonic weighted average) and exploration of this possibility is a subject in itself, and so we leave it for the future research.

The second assumption is that all DMUs face the same input prices. This type of "Law of One Price" assumption might be viewed as somewhat undesirable (see Kuosmanen et al. (2006) for a related discussion), rarely fulfilled in practice, yet it is a necessary condition to obtain (3.7). Noteworthy, it is not the first time that a positive aggregation results in economics require additional, often very strict or even undesirable assumptions (e.g., recall about the aggregation of demands over goods or consumers). One

[1] We thank anonymous referee for this important insight.

could think of it as a theoretical abstraction, a simplification consistent with many existing economic models (e.g., assumption on perfectly competitive input markets, etc.) as well as in empirical analysis. One could also think of this assumption here as a tool to get a theoretically justifiable benchmark for aggregate scale measures that accounts for economic importance of each DMU in the sample. The question of deviation from this benchmark due to price variability across firms might be another direction of research.

The third assumption is that each DMU individually as well as a group of them exhibit *optimizing behavior* (minimization of cost functions for this case). This type of assumption is usual in neo-classical economics and, in a sense, not only restricts but also empowers the meaning of the result of derivations.

Overall, anyone who would or did use the aggregation scheme (3.10)—(3.11), however intuitive it may sound, must essentially accept these three assumptions, whether they like them or not. Alternative route would be to show that some other assumptions would imply this or other aggregation scheme and we hope our work would encourage such research.

Finally, it is worth noting that this paper is just the first step in laying the new theoretical foundation for analyzing aggregate scale elasticities. The next step must be laying out the statistical foundation, because besides presenting an average of the data, researchers often need to do some statistical testing for inferring on various hypotheses. A hypothesis of an interest, for example, might be whether the true aggregate scale elasticity for a particular *sub-group* is different from, say, unity (i.e., local CRS) or not. Another hypothesis of interest might be whether the scale elasticity is equal across different sub-groups or not, or across time for the same sub-groups, etc. For the context of using the equally-weighted sample mean this issue is somewhat standard. On the other hand, estimating characteristics of the sampling distribution of a statistic for a *weighted* mean and related testing is more complicated. One potential and currently popular solution is to adapt the bootstrapping techniques (e.g., see Efron, 1979).

This extension is likely to be similar in some parts to the recent work of Simar and Zelenyuk (2007) for bootstrapping the aggregate efficiency scores, but is a subject in itself and so we leave it for future research.

6. Concluding Remarks

In this paper we introduced the *aggregate* primal and dual measures of scale elasticities and showed under what assumptions they can be obtained by weighted aggregation of *individual* scale elasticities. Specifically, we showed that when economic agents follow cost minimizing behavior, aggregate technology is characterized by the summation of individual input requirement sets and all individuals face the same input prices, then the observed cost shares would be theoretically justifies weights in the arithmetic average of the individual scale elasticities. We also suggested price independent weights—when dual (price or cost) information is not available. Similar analysis can be applied to the context of the revenue and the profit maximizations, in which case the output distance function and the directional distance functions (or, alternatively, the hyperbolic measures), respectively, might be used to characterize technology.

The aggregation results we derived can also be extended to the case of aggregation *within* distinct *sub-groups* (e.g., public and private, regulated and non-regulated, etc.) and then for aggregation *between* these sub-groups into a larger group (for the related theorem, see Simar and Zelenyuk (2007)). Similar derivations can be done for the case when revenue function or output distance function is used to measure scale elasticity—the approach useful when cross-DMU reallocation of output must be allowed. A separate research is needed to derive aggregation scheme for the case when cross-DMU reallocation of both inputs and outputs must be allowed, e.g., as in Nesterenko and Zelenyuk (2007).

Finally, similar analysis can be applied for other 'derivatives' of the cost, revenue and profit functions—to obtain aggregation of demand and supply functions (using Shephard's/Hotelling lemmas), which we have not focused on in this paper, but it would be a natural direction for future research. Furthermore, bootstrapping procedures must be developed for the benefit of the practical use of these measures for enabling various statistical hypotheses testing.

Overall, we believe the aggregate measures of scale elasticity that were presented and theoretically justified in this paper, must serve as useful tools for empirical researchers in analyzing and presenting their results as well as for laying the new grounds for future theoretical work.

References

Efron, B. (1979) "Bootstrap methods: another look at the jackknife," *Annals of Statistics* 7, 1-26.

Färe, R., S. Grosskopf and V. Zelenyuk (2004) "Aggregation of Cost Efficiency Indicators and Indexes Across Firms," *Academia Economic Papers* 32:2, 395-411.

Färe, R. and D. Primont (1995) *Multi-Output Production and Duality: Theory and Applications*, Kluwer Academic Publishers, Boston.

Färe, R. and V. Zelenyuk (2003) "On Aggregate Farrell Efficiency Scores," *European Journal of Operations Research* 146:3, 615-620.

Färe, R. and V. Zelenyuk (2007) "Extending Färe and Zelenyuk," *European Journal of Operational Research* 179:2, 594-595.

Koopmans, T.C. (1957) *Three Essays on the State of Economic Analysis*, McGraw-Hill, New York.

Kuosmanen, T., L. Cherchye, and T. Sipiläinen (2006) "The Law of One Price in Data Envelopment Analysis: Restricting Weight Flexibility Across Firms," *European Journal of Operational Research* 170:3, 735-757.

Li, S.-K., Ng, Y.-C., (1995) "Measuring the Productive Efficiency of a Group of Firms," *International Advances in Economic Research* 1:4 377-390.

Nesterenko, V. and V. Zelenyuk (2007) "Measuring Potential Gains from Reallocation of Resources," *Journal of Productivity Analysis* 28:1, 107-116.

Shephard (1970) *Theory of Cost and Production Function*s, Princeton: Princeton University Press.

Simar and Zelenyuk (2007) "Statistical Inference for Aggregates of Farrell-type Efficiencies," *Journal of Applied Econometrics* 22:7, 1367-1394.

Zelenyuk, V. (2006) "On Aggregation of Malmquist Productivity Indexes," *European Journal of Operational Research* 174:2, 1076-1086.

Zelenyuk, V. (2010) "Aggregation of Economic Growth Rates and of its Sources," CEPA Working Papers Series WP032010, School of Economics, University of Queensland, Australia (forthcoming in *European Journal of Operational Research*).